Applying Toponome Imaging System for studying Colon Cancer.

Mr. Sayantan Bhattacharya

MBBS, MRCS, PGA (Medical Education)

A Thesis submitted to the University of Warwick for the Degree of Doctor of Medicine.

Life Sciences

University of Warwick

Coventry, CV4 7AL

July, 2011

Page

A. Contents

CHAPTER 1

CHAPTER 2

B. List of Figures

Figure 1: This figure demonstrates the sequence of events leading to colon cancer from normal colon. In figure 1A, the sequence taking place in Familial adenomatous polyposis syndrome and majority sporadic cases is shown. The sequence depicted in Figure 1B, demonstrates the events in hereditary non-polyposis colorectal cancer syndrome and remaining sporadic cases.

Figure 2: This figure demonstrates the canonical Wnt / β-catenin signalling pathway. Under normal circumstances, (Figure 2A) GSK3-β phosphorylates β-catenin, inducing the later to undergo proteasomal degradation. However, if this phosphorylation is inhibited (Figure 2B) by Wnt signalling pathway (via activated frizzled and dishevelled respectively), β-catenin translocates in the nucleus, replaces protein Groucho (from Groucho-TCF4 complex) to form a complex with TCF-4. This molecular complex is an active transcription factor and is responsible for inducing several proteins related to cell proliferation.

Figure 3: This figure demonstrates the mechanisms by which K-ras activates several transcription factors like AP1, elk1, c-jun – c-fos complex and c-myc to promote cell proliferation through activation of Raf-Kinase and inhibits apoptosis, prevents proteasomal degradation of β-catenin and induces proteins related to cell proliferation through activation of akt-kinase.

Figure 4: This figure demonstrates the anti-carcinogenic roles of TGF-β in colon cancer by promoting cell cycle arrest at G1 phase and apoptosis.

Figure 5: This figure demonstrates mechanisms related to the regulation of p53 and its anti carcinogenic role by inhibiting anti- apoptotic factors (bcl-2, surviving and bcl-xn) and promoting proteasomal degradation of β-catenin.

Figure 6A: Mechanisms related to Non-steroidal anti-inflammatory drugs mediating 15-lipoxygenase-1 induced apoptosis are demonstrated in this figure.

Figure 6B: This figure demonstrates the effects of PPAR and 15-LOX-1 in colon cancer.

Figure 6C: This figure shows 15-lipoxygenase induced apoptosis dependent on p53, p21 and NAG-1.

Figure 7: This figure demonstrates the toponome imaging system used for performing these experiments with colon cancer and normal colon. Numbered arrows are used to identify the important components:

1. Buffer tray covered by thin aluminium foil, to be pierced by the pipette tip.

2. Antibody tray, to be filled with antibody solutions of specified dilutions.

3. Pipette tip attached to the end of the robot arm.

4. Camera attached to the side port.

5. Microscope

6. Glass slide for the tissue section.

7. Touch panel control for microscope.

Figure 8: Figures showing results from optimisation runs. Figure 8a shows a section of colon cancer where antibodies did not localise any proteins due to low concentration during optimisation. Figure 8b shows complete white out image due to much higher concentration of antibody Ki67 used on a colon cancer specimen. Figure 8c shows an artefact with false fluorescence. Of note, these artefacts can result in false staining. Figure 8d demonstrates unopened shutter, a problem encountered during TIS runs in Warwick. Figure 8e shows a blurred phase image, possible causes being unco-ordinated shutter opening or mal-positioned condenser of the microscope or moving stage while imaging process in going on. Figure 8f shows artefact and blurring of phase image due to auto-focussing on the artefact rather than on the tissue section

Figure 9: Phase (unstained) and OF (post fluorescent) images from the 21 tags used on cancer tissue. The last image is of an adjacent section from the same cancer tissue block stained with routine H&E to demonstrate cancer in the tissue.

Figure	Antibody/Tag/Description	Figure	Antibody/Tag/Description
a	Phase image	m	Cyclin A
b	PBS, post bleaching	n	CEA
c	DAPI	o	CD166
d	Laminin	p	PCNA
e	CK-19	q	CD44
f	NCAM	r	Muc-2
g	CD36	s	Muc-1
h	CK-20	t	CD29
i	Ki67	u	CD57
j	Bax	v	CD133
k	Bcl-2	w	EpCAM
l	Cyclin D1	x	H&E image of adjacent section from same cancer colon tissue block.

Figure 10: Phase and OF (post fluorescent) images of normal tissue. The last image is of an adjacent section from the same normal tissue block stained with routine H&E to demonstrate absence of cancer in the tissue.

Figure	Antibody/Tag/Description	Figure	Antibody/Tag/Description
a	Phase image	m	Cyclin A
b	PBS, post bleaching	n	CEA
c	DAPI	o	CD166
d	Laminin	p	PCNA
e	CK-19	q	CD44
f	NCAM	r	Muc-2
g	CD36	s	Muc-1

h	CK-20	t	CD29
i	Ki67	u	CD57
j	Bax	v	CD133
k	Bcl-2	w	EpCAM
l	Cyclin D1	x	H&E image of adjacent section from same normal colon tissue block.

Figure 11: Glandular structure in: colon cancer (figure a, phase contrast and b, DAPI) and normal colon: (figure c, phase contrast and d, DAPI). The cell cytoskeleton, marked by CK-19 and CK-20, is disrupted in colon cancer (figures a and b) and clumped in the perinuclear area (figure b). In contrast, in normal colon the cell cytoskeleton is much more evenly distributed (figure c and d).

Figure 12: Subset of colon cancer cells identified by CMPs. Cells marked by CMPs with lead proteins; CD133, CD44, EpCAM and CD166 in the top row (a and b) and cells marked by all the CMPs with CD166, CD44 and EpCAM as lead proteins in the middle row (c and d) in cancer tissue. The CMPs in the top row have been shown by black arrows and those in the middle row by white arrows. The bottom row (figure e, f and g) shows magnified views of all CMPs from figure c. All the CMPs in the top and middle sections have CD133 co-expressed along with CD166, CD44 and EpCAM, except the single cell (CMP=colour purple) towards the centre of the middle row of images (c and d). This cell is magnified in figure g in the bottom row.

Figure 13: Schematic diagram indicating CMP motifs specific for glandular cells and stromal cells in colon cancer (in 13A) and normal colon (in 13B). The proteins are indicated by numbers as in the table in the bottom right hand corner. The inner circle shows proteins (highlighted in red) which are lead proteins. The outer circle shows the

other proteins in the motif, the ones highlighted in yellow are anti-colocalised or absent and the rest are variably present (wild card proteins).

Figure 14: Illustration of the 15 most frequent CMPs which are unique to normal colon (not seen in the cancerous part of the same colon).

Figure 15: Illustration of the 15 most frequent CMPs which are unique to colon cancer (not seen in the normal part of the same colon).

In Figures 14 and 15, each colour represents one specific CMP with subcellular resolution. The left panel shows CMPs superimposed on the phase contrast image; the right panel shows the same CMPs superimposed on the corresponding DAPI fluorescence image showing the location of cellular nuclei (white). Note the CMP coding list shown on the bottom of this figure with (i) specification of molecules co-mapped by TIS (top line no.´s 1 – 21 of the list), (ii) colour specification of CMPs, in which every CMP expresses a unique combinatorial binary code (1 bit information per protein/molecule) corresponding to the colours in the phase- and DAPI- images; (iii) CMP motifs (bottom line of the list). This CMP motif represents the higher order organisation of the 15 different CMPs as a group with features common to all CMPs described by a three symbol code[1]: **L**, lead protein/molecule (present in all CMPs); **A** (proteins/molecules which are absent in all CMPs = anticolocated); **W** (wild card proteins/molecules variably associated with the L and A molecules of CMPs). Bar: 10 µm.

Figure 16: **1st row;** All 34 Cancer unique CMPs in phase (a) and DAPI (b) images, with following lead proteins: CD133, CD166, EpCAM, CK-20 and Muc-1. Of note, these cells also co-express Ki67, PCNA, Bax, Bcl-2 and Cyclin-A. **2nd row;** CMPs in 1st row images are magnified in phase (c) and DAPI (d) images.

C. List of Tables

Table 1: List of all the antibodies with their locus link annotations, used in their order in the original experiments.

Table 2: List of the 34 “unique” CMPs present in cancer with following lead proteins: CD133, CD166, EpCAM, Muc-1, Laminin, CD57 and CK-20. The cells marked by these CMPs also expressed Ki67, PCNA, Bax, Bcl-2 and Cyclin-A. Respective colour codes are shown by the sides of each CMP. Of note, CD133, CD166 and EpCAM are putative colon cancer stem cell markers.

Table 3: List of 25 most frequent CMPs found in stromal cells in colon cancer with individual colour codes. Of note, the lead proteins **(L)** are CD57, Laminin and PCNA. Proteins CD29, CD36, CEA, CK-19, Muc2, Muc1 and NCAM were found to be **(A)** anti-co localised and rest of the proteins were wild cards **(W)** and hence variably expressed. A schematic diagram is provided in Figure 13 to illustrate this.

Table 4: List of 29 most frequent CMPs found in neoplastic glandular cells in colon cancer with individual colour codes. Of note, the lead proteins are CEA, EpCAM, CK20 and Ki67. Proteins CD36 and CD44 were found to be anti-co localised and rest of the proteins were wild cards and hence variably expressed. A schematic diagram is provided in Figure 13 to illustrate this.

Supplementary Tables:

Supplementary Table 1: Complete list of all CMPs in cancer colon. Each horizontal row represents a CMP. The top row contains the names of the antibody tags used and the last column on the right depicts the relative frequencies of the respective CMPs. Total CMPs in this list = 6813

Supplementary Table 2: Complete list of all CMPs in normal colon. Each horizontal row represents a CMP. The top row contains the names of the antibody tags used and the

last column on the right depicts the relative frequencies of the respective CMPs. Total CMPs in this list = 32009. Of note, the list is almost five times that of cancer.

Supplementary Table 3: Complete list of all CMPs unique in cancer. Each horizontal row represents a CMP. The top row contains the names of the antibody tags used and the last column on the right depicts the relative frequencies of the respective CMPs. Total CMPs in this list = 5708.

D. Acknowledgements

I take this opportunity to thank Dr. Michael Khan, my supervisor, who supported my endeavours to successfully perform research with this novel technology. I would also like to thank my clinical supervisor, Mr George Mathew, who was instrumental in setting up this project, arranging funds, providing administrative back-up and helping with tissue collection in the hospital. Apart from this ground work, he was my moral support, believing in me during the highs and lows that I experienced during research.

I would like to thank Professor David Epstein for his support in helping me with data analyses.

A significant contribution in this project was from Professor Walter Schubert (Magdeburg), who was the main proponent of TIS. I spent two weeks at his laboratory to learn the basic principles. I would therefore like to convey my sincere gratitude to him and his team members; Andreas Krusche, Reyk Hillert, Anne Giesler and Doris for helping me grasp this technology and perform sections.

I am thankful to Dr. Ernie Ruban, consultant Histo-pathologist, George Eliot Hospital, for his contribution for selection of relevant visual fields for this study.

I would also like to thank Dr. Sylvie Abouna and Dr. Sevi Zervou for their support as team members and helping me with learning the basic methods of immunohistochemistry, tissue sectioning and staining techniques.

I greatly acknowledge the help provided by Dr. Vinod Patel, Chairman, DARE (Department of Audit and Research) committee, George Eliot Hospital with respect to funding this project.

Technological and biological toponome studies were supported by BMBF grants Biochance, CELLECT, NBL3, NGFN2, and NGFN plus, as well through DFGschu627/10-1, the Innovationskolleg INK15, and through the Klaus Tschira foundation, as well as through the LOM project (Med Fak, Univ of Magdeburg, Germany).

I dedicate this work to my wife, Miami, without whose love and support none of this would have been possible.

E. Declaration

I declare that the results provided in this thesis were conducted by the author under supervision of Dr. Michael Khan (Warwick University), Mr. George Mathew (George Eliot Hospital), Professor Walter Schubert (Magdeburg) and Professor David Epstein (Warwick University). None of these results were used in any previous application for a degree.

The colon tissue(s) used in this study for optimisation of TIS or obtaining results were obtained from colectomy specimen of surgical patient at George Eliot Hospital after proper consent and were in accordance with Ethics Committee approval obtained.

Sources of information have been specifically acknowledged by reference.

F. Summary

Introduction

In a proof of principle study, we have applied an automated fluorescence toponome imaging system (TIS) in order to examine, whether TIS can find protein network structures distinguishing cancerous from normal colon tissue from same patient.

Methods/Materials

Cancer specimen and corresponding normal tissue were harvested at colectomy from a single patient. 5μm sections were then prepared for TIS using a battery of different antibodies, including a number of putative CSC markers. Expression of multiple protein clusters was determined and Combinatorial Molecular Phenotypes (CMPs) were analysed, using specific image-analysis tools.

Results

By using a three symbol code and a power of combinatorial molecular discrimination (PCMD) of 2^{21} per sub-cellular data point in one single tissue section, we demonstrate an *in situ* protein network structure, visualized as a mosaic of 6,813 protein clusters (Combinatorial molecular phenotype or CMPs) in the cancerous part of the colon. By contrast, in the histologically normal colon, TIS identifies nearly 5 times the number of protein clusters as compared to the cancerous part (32,009).

Discussion and Conclusion

By sub-cellular visualization procedures we found, that many cell surface membrane molecules were closely associated with the cell cytoskeleton as unique CMPs in the normal part of the colon, while the same molecules were disassembled in the cancerous part, suggesting presence of dysfunctional cytoskeleton-membrane complexes. As expected, glandular and stromal cell signatures were found, but interestingly also potentially TIS signatures identifying a very restricted subset of cells expressing several putative stem cell markers, all restricted to the cancerous tissue. The detection of these signatures is based on the extreme searching depth, high degree of dimensionality, and

sub-cellular resolution capacity of TIS. These findings provide the technological rationale for the feasibility of a complete colon cancer toponome to be established by massive parallel high throughput/high content TIS mapping.

G. Abbreviation

2-DE	2-Dimensional Gel Electrophoresis
2-DL	2-Dimensional Liquid Chromatography
2D-PAGE	2-Dimensional denaturing Polyacrylamide Gel Electrophoresis
APC	Adenomatous Polyposis Coli
APOCON	Aminopeptidase-N Polarization Control Network
ARF	Alternative Reading Frame
BANK-1	B-cell Scaffold Protein with Ankyrin Repeats-1
Bax	Bcl-2 Associated X-Protein
Bcl-2	B-Cell Lymphoma-2 Protein
BCNP-1	B-Cell Novel Protein-1
BMP	Bone Morphogenetic Protein
CA19-9	Carbohydrate Antigen 19-9
CAGE	Cap Analysis of Gene Expression
CAM	Cell Adhesion Molecule
CBC cells	Crypt Base Columnar Cells
CD	Cluster of Differentiation
CDK	Cyclin Dependent Kinase
CDS	Cytidine Deaminase
CE	Capillary Electrophoresis
CEA	Carcino Embryonic Antigen
CK	Caseine Kinase
CK	Cytokeratin
CMP	Combinatorial Molecular Phenotype
CPP	Combinatorial Protein Pattern
CRMP-2	Collapsing Response Mediator Protein-2

CSC	Colon cancer Stem Cell
DAPI	Diamino Phenyl Indole
DiGE	Difference Gel Electrophoresis
DNA	Deoxy-Ribonucleic Acid
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial Mesenchymal Transitions
EpCAM	Epithelial Cell Adhesion Molecule
ER	Estrogen Receptor
ESI	Electro Spray Ionization
FAP	Familial Adenomatous Polyposis
FGFR-3	Fibroblast growth factor receptor 3
FITC	Fluoresceine-iso-thio-cyanate
FOBT	Faecal Occult Blood Testing
G-CSF	Granulocyte- Colony Stimulating Factor
GSK	Glycogen Synthase Kinase
GWAS	Genome Wide Association Studies
H&E	Haematoxylin and Eosin
HCAM	Homing Cell Adhesion Molecule
HER	Human Epidermal Growth Factor Receptor
HIF	Heat Inducible Factor
HNPCC	Hereditary Non-Polyposis Colorectal Cancer Syndrome
HODE	HydroxyOctaDecadienoic Acid
HUPO	Human Proteome Organization
IEF	Iso Electric Focussing
ILK	Integrin Linked Kinase
IPG	Immobilized pH Gradient
JPS	Juvenile Polyposis Syndrome

K-RAS	Kirsten Rat Sarcoma viral oncogene homolog
LC	Liquid Chromatography
LGR-5	Leucine rich repeat containing G Protein coupled Receptor
LIT	Linear Ion Trap
LOX	Lipoxygenase
LRP-5	Low density lipoprotein Receptor related Protein -5
MALDI	Matrix Assisted Laser Desorption / Ionization
MAPK	Mitogen Activated Protein Kinase
MCSF	Macrophage- Colony Stimulating Factor
mi-RNA	micro-Ribonucleic Acid
MLH-1	MutL Homolog-1
MMP	Matrix Metalloproteinase
MMR	MisMatch Repair
MoPPi	Modular Processing Pipeline
MPSS	Massively Parallel Signature Sequencing
m-RNA	messenger-Ribonucleic Acid
MS	Mass Spectometry
MS4A-1	Membrane Spanning 4 domains, subfamily A member 1
MSH2	MutS Homolog-2
MSI	MicroSatellite Instability
MSS	MicroSatellite Stable
m-TOR	mammalian Target Of Rapamycin
Muc	Mucin
NCAM	Neural Cell Adhesion Molecule
nc-RNA	non coding-Ribonucleic Acid
NIH	National Institute of Health
NNMT	Nicotinamide N-Methyl Transferase

NOD	Non Diabetic
NSAIDs	Non Steroidal Anti Inflammatory Drugs
PBS	Phosphate Buffer Solution
PCMD	Power of Combinatorial Molecular Discrimination
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PISA	Protein in situ Array
PLC	Phospholipase C
PMS-2	Post Meiotic Segregation-2
PPAR	Peroxisome Proliferator Activated Receptor
PSA	Prostate Specific Antigen
PSME-3	Proteasomal Activator Complex subunit
PTEN	Phosphatase and Tensin Homologue
RIIBC	Repeated Incubation Imaging Bleaching Cycles
RNA-seq	RNA- sequencing
SAGE	Serial Analysis of Gene Expression
SAMP	Serine-Alanine-Methionine-Proline
SCID	Severe Combined Immunodeficient
SEK	Simple Epithelial Keratin
SELDI	Surface-enhanced Laser Desorption/Ionization
SNP	Single Nucleotide Polymorphism
SSRs	Simple Sequence Repeats
TA cells	Transiently Amplifying Cells
TACSTD-1	Tumor-associated Calcium Signal Transducer 1
TCF-4	T-Cell Factor-4
TGF	Transforming Growth Factor
TIMP-1	Tissue Inhibitor of Metalloproteinase Type 1

TIS	Toponome Imaging System
TOF	Time of Flight
t-RNA	transfer-Ribonucleic Acid
UPS	Ubiquitin Proteasomal System

H. Definitions:

Toponome: Toponome is derived from the Greek word Topos (Place) nomos (Law). It is defined as the entirety of protein network in a cell, the final phase of protein interaction, deciding cellular characteristics.

CMP: Combinatorial Molecular Phenotype, also referred as Combinatorial Protein Pattern (CPP), is the smallest unit of Toponome. Each such CMP represents a combination or cluster of proteins per data point in the visual field.

CMP motif: A combination of CMPs which share any unique character, for example present or absent in a particular state (normal versus cancer) in a particular structure (glandular structures versus stromal cells), forms a motif.

Lead protein: In a CMP motif, the protein(s) which are present in all the CMPs is / are termed as Lead Protein(s).

Anticolocalised protein: In a CMP motif, the protein(s) which are absent in all the CMPs is / are termed as Anti-colocalised Protein(s).

Wild Card protein: In a CMP motif, most of the proteins are present in some CMPs but absent in others. These variably expressed proteins are termed as Wild Card Proteins.

Three symbol code: A three symbol code is a way to describe a CMP motif by its Lead Protein(s), Anticolocalised Protein(s) and Wild Card proteins, thus represented by letters L, A and W respectively.

CHAPTER 1

1.1. INTRODUCTION

Colorectal cancer is responsible for a major proportion of cancer related mortality in the western world (second commonest in England and Wales and third in US). Over 30,000 new cases of colon cancer are diagnosed each year, more than 25% of which are Dukes' C (advanced) at presentation[1]. Colon cancer originates from the colonic epithelial cells undergoing sequential mutations. These putative mutations are responsible for the transition of normal epithelial cells to adenomas with increasing dysplasia, finally resulting in colorectal adenocarcinomas, as described by adenoma-carcinoma sequence (shown in Figure 1). Such mutations could either be a result of inborn genetic aberrations like Familial adenomatous polyposis (FAP) or could be due to environmental carcinogens like tobacco smoke or chronic inflammatory states involving the colon and rectum[2].

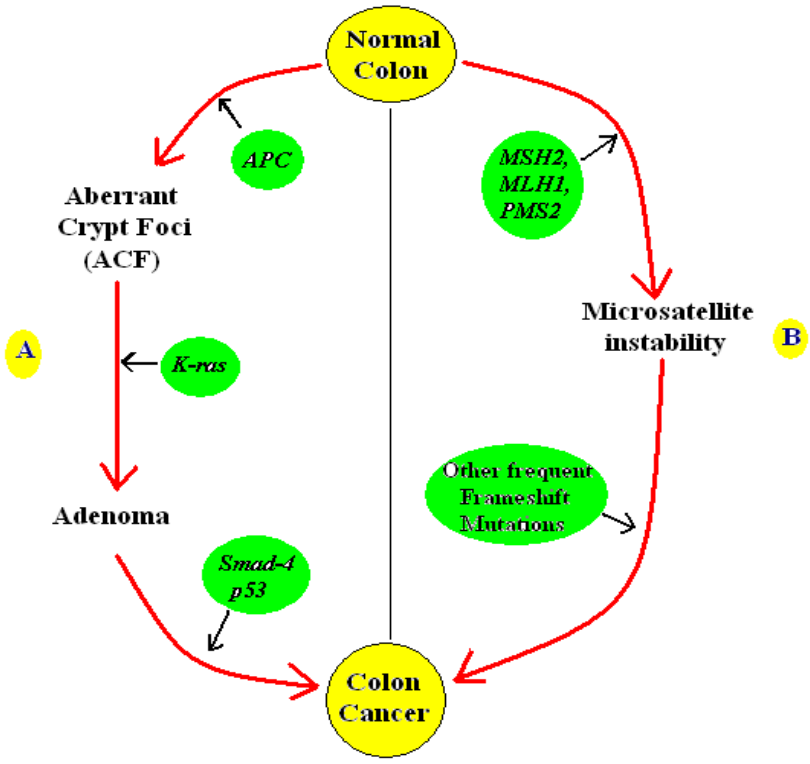


Figure 1: This figure demonstrates the sequence of events leading to colon cancer from normal colon. In figure 1A, the sequence taking place in Familial adenomatous polyposis syndrome and majority sporadic cases is shown. The sequence depicted in Figure 1B,

demonstrates the events in hereditary non-polyposis colorectal cancer syndrome and remaining sporadic cases. *(For further details please refer to Voutsadakis I. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2* [3]*.)*

Origin of colorectal cancer is a hotly debated topic and is explained by two models:

- The stochastic or clonal evolution model
- The cancer stem cell model.

In the clonal evolution model, most of the cells in the tumour are thought to possess the capacity of self-renewal and contribute to the growth of the tumour. The differentiation, intra clonal genetic and epigenetic variations and other micro environmental influences are responsible for the tumour heterogeneity in this model.

The cancer stem cell model rests on the hypothesis that, some special "stem cells" in the colonic epithelium possess the unique capacity of differentiation and producing a clone of cancer cells. These newly formed clone of cancer cells lack the potential for self-renewal. Along with these cancer cells, the stem cells also produce new cancer stem cells, which are responsible for further propagation of the cancer. According to this theory, aberrant differentiation from stem cells contributes to the heterogeneous nature of the tumour[4].

1.1.1. Adult intestinal stem cells and cancer stem cells:

Adult Intestinal Stem cells:

Stem cells, by definition, are capable of reconstituting a tissue, autonomously in its fundamental morphological and functional aspects. In adult human gut, epithelium completely self renews itself within 5 days, producing more than 10^{10} differentiated cells every day[2]. Upon activation, intestinal stem cells undergo a specific type of mitosis called "asymmetric cell division", to produce two cells of different developmental potential: one daughter stem cell, capable of further stem cell activity and one non-stem cell, which continue to replicate, migrate upwards in the crypt – villous axis and

differentiate into mature cells[5,6]. These first generation non-stem daughter cells produced by the stem cells are known as Transiently Amplifying cells (TA cells). Under normal circumstances, TA cells divide actively for a fixed length of time, to produce a cellular pool, which then differentiates to form a tissue[7,8]. The intestinal stem cells and their TA cell progeny are located towards the base of the crypts of Lieuberkuhn of the intestine. As these cells migrate towards the surface (except the Paneth cells, which stay at the bottom of the crypts), they cease to proliferate and start differentiating into 5 major cell lineages[9,10]:

- Absorptive enterocytes
- Mucin producing goblets
- Peptide hormone producing neuroendocrine cells
- M cells involved with antigen transport from gut lumen to Peyer's patches.
- Microbicidal Paneth cells.

A normal human crypt is usually quite small, contains multiple stem cells and composed of nearly 2000 cells originating from different stem cells. The exact location of stem cells in a crypt is still a matter of debate, the major obstacle being the absence of unique molecular markers. Thus assumed functional properties like radio-sensitivity and label retention were used for their identification. It has been proposed that in the intestine, the stem cells reside in a protective "niche" or the "+4 position", the niche playing an important role in maintaining these cells in their immature state. If removed from their niche, the respective stem cells are thought to lose their multi-potent nature. These +4 position cells were found to be both radiosensitive and label retaining, with elevated expression of *Bmi1* (mouse models), a gene thought to be needed for maintaining stem cells[11-13]. Alternatively, mouse models were used by two scientific groups to identify small epithelial cells interspersed in between Paneth cells in this niche, thought to have stem cell properties. These cells were named Crypt Base Columnar (CBC) cells[14,15]. Though the evidence is poor, a fenestrated sheath of intestinal sub-epithelial myo-

fibroblasts, surrounding the crypts is thought to form this “protective niche” for the intestinal stem cells. These myo-fibroblasts extend throughout the lamina propria and merge with the layer of pericytes, surrounding the blood vessels[16]. Wnt pathway is one of the main signalling pathways that direct stem cell proliferation. In presence of *in vivo* Wnt signalling (please refer to Figure 2 and detailed discussion under morphogenetic pathways section), β-catenin phosphorylation is inhibited and intra nuclear active transcription factor complex consisting of TCF-4 and β-catenin results. The activity of the Wnt pathway, in conjunction with other relevant pathways like BMP and Notch, is vital for stem cell proliferation and the proper organisation of the intestinal epithelium. A Wnt signalling gradient apparently exists along the intestinal crypt villous axis. The highest level of influence of this signalling pathway is at the base of the crypts. As a proliferating cell migrates away from the base, it progressively loses its proliferating nature and starts differentiation. This explains why the stem cells are mostly found at the bottom of the crypts. Likewise, the transcriptional target proteins of this pathway were expressed abundantly in the crypts. Wnt signalling is also essential for the expression of cell surface receptors and corresponding ligands involved in the organisation of the epithelium, for example, EphB / EphrinB receptors and ligand system that dictates the position of different cells along the crypt-villous axis[17]. Potten and colleagues found that the upper and lower limits of cell proliferation in the colorectal mucosa are expressed by the 95^{th} and 5^{th} percentiles at cell positions 43 and 4 respectively and the peak labelling index was seen at cell position 15[18]. Musashi-1, a RNA-binding protein has been considered as an intestinal stem cell marker in the past[19]. Cells expressing Musashi-1, in their cytoplasm and nucleus, were demonstrated in abundance below position 10, with a peak expression at positions 3-4, by Nishimura *et al*. Based on their findings, the authors suggested that the location of intestinal stem cells in humans is between positions 1 – 10 in the crypt – villous axis[20]. Though the exact role of Musashi-1 in human intestinal stem cells is not yet clearly understood, it has been shown to interact with Notch

signalling pathway and contribute towards maintaining the stem cells in an immature state in some mammalian tissues. Presumably, it plays a similar role in the human intestine[21]. Among a small group of genes, whose presence was restricted mainly in the crypt base, *LGR5* (Leucine rich repeat containing G protein coupled receptor) encoded for a G-protein coupled receptor and found mainly in CBC cells. Lineage trace *in vivo* experiments with mouse models by Barker and colleagues confirmed the same. Furthermore, it was also shown by the authors that these cells were able to divide approximately once per day and thus generate all cell types of the intestine for the lifetime of the mouse[22]. Recently, CBC cells were also found to express high levels of CD133 as well, and lineage trace experiments with CD133 also indicated that CBC cells are stem cells[23]. Beta-1 integrin subunit[24], EphB receptors[25] have also been considered as markers for intestinal stem cell.

Cancer Stem cells:

A tumour can be considered as an “aberrant organ” that has its origin from a transformed stem cell (monoclonal), producing a progeny of daughter cells, which become heterogeneous due to subsequent differentiation. This heterogeneity is both in terms of morphological and functional characteristics of the tumour cells. Functionally, cells in a tumour have different proliferating potentials. Interestingly, *in vivo* and *in vitro* experiments have shown that only a minority of the cells in a tumour mass are tumorigenic and only a few of these possess high proliferating capacity[26,27]. As previously discussed, origin of colon cancer has been explained by 2 models, the stochastic or clonal evolution model and that of the cancer stem cell. Previously, tumour origin was thought to result according to the principles of the stochastic model. In this model, all cells in the tumour were considered to have equal ability to divide and sustain the tumour. The morphological heterogeneity in tumours was explained in this model by genetic instability of individual cells, causing them to follow different pathways of aberrant differentiation. These changes were also held responsible for explaining the cell

intrinsic and extrinsic processes associated with tumour development, such as angiogenesis or Epithelial Mesenchymal Transitions (EMT). The therapeutic consequence in this model was that, in order to successfully treat a tumour, all of its cells needed to be removed or destroyed, as all of them were equally capable of generating a new tumour and cause cancer relapse. Over the last two decades, with the introduction of stem cell model, the tumour is studied applying the principles of stem cell biology. This concept has been hugely supported by reports of small subsets of specific cancer cells, named as cancer stem cells (CSC), having the capacity of self-renewal and potential to proliferate for an indefinite time, when grafted in immuno-compromised animal models. These cells might have originated from the normal stem cells of the organ concerned or the transiently amplifying (TA) cells, as a result of putative mutations in either of them[28]. There is also some suggestion that the cell of origin may also be a late progenitor cell or even an early differentiated cell[29]. According to this "cancer stem cell" concept, tumour cells are thought to maintain a hierarchical organization consisting of a population of self renewing stem cells at the bottom of the hierarchy which undergo atypical cell division to produce more stem cells on one hand, thus renewing its progeny along with progenitor cells with limited proliferative capacity and differentiated cells with no proliferative potential on the other hand. Quite interestingly, the clinical implication of this model is that by removing the stem cell the tumour will no longer be capable of growing and therefore can be downgraded. Henceforth, future therapeutic modulation can be targeted against specific stem cells and not reducing or debulking the entire tumour[30]. However, controversies still exist in this field as inability of most other colon cancer cells in replicating in a xenograft model might reflect their limitations in making the correct contacts, etc, in the ectopic new host or even the level of immune suppression in the host animal (for example, mouse) model[4,31]. Mutations in Wnt pathway, either inactivating tumour suppressor functions of *APC* or *Axin2* or activating oncogene *β-catenin* are considered to be responsible for the vast majority of colon

cancers. Under both circumstances, β-catenin can be localized in the nucleus (Figure 2, morphogenetic pathways) and Wnt target genes are found to be activated[32]. As shown in Figure 1, after Wnt pathway, carcinogenic mutations in *KRAS*, *SMAD4* and *p53* pathways are responsible for conversion of a normal colonic cell to a cancerous cell through "adenoma – carcinoma sequence" (Figure 1, please see above)[33]. Along with these putative mutations in the process of carcinogenesis, additional genetic alterations are also known to happen, like those leading to loss of positional clues of cancer cells, leading to disorganized cellular arrangement. From our current understanding, un-mutated Wnt pathway (Wnt gradient, discussed above) is responsible for establishing the crypt – villous border. It is this same Wnt gradient which controls expression of EphB/EphrinB receptor and ligands responsible for positional migration of the intestinal cells, with proliferating cells at the bottom of the crypt and more differentiated cells towards the top in the villous. A disruption of this system with mutations in Wnt pathway is associated with EphB2 down regulation, a step found to be required for cells in an adenoma to progress to carcinomatous cells, freeing themselves of their positional constrains[34]. Studies using CRE inducible activation of Wnt pathway were used to demonstrate that the cell of origin of adenomas was none other than the mutated intestinal stem cell. In either of these studies CRE was expressed from the endogenous *LGR5* or *CD133* locus of the stem cell marker genes and both LGR5 and CD133 proteins were found to be effective inducers of Wnt pathway leading to efficient tumour formation[23,35]. However, not all cancer cells in an APC induced tumour had equal Wnt pathway activity[32,36]. Again, loss of function mutation of *APC* in progenitors and differentiated cells in mouse models rarely induced tumour formation[35]. Two scientific groups reported that colon cancer cells expressing CD133 were uniquely capable of initiating, proliferation and sustaining the tumour in NOD/SCID mice and therefore claimed that CD133 is a marker for cancer stem cell in colon[37,38]. Later a subset of human colon cancer cells expressing CD166, CD44 and EpCAM were also found to have similar properties of producing tumours in mice models and hence these proteins

were considered by the authors as additional markers for colon cancer stem cell[39]. In both these cases, the new tumours generated in mice exactly phenocopied the human colorectal cancers from which the parent cells were isolated. Additional support for stem cell model came from the observation that these cancer stem cells were isolated from metastases of human colon cancer. The considerable difference in frequency of stem cells in between studies can be explained by different antibodies recognizing different forms of marker proteins used (for example glycosylated CD133). Although, CD133 was previously thought to be a specific marker of colon cancer stem cells[23], it was later shown to be expressed in many other intestinal cells including all normal stem cells and progenitor cells[40,41]. Another interesting finding that has recently come to light is that the stemness of a cancer stem cell is greatly controlled by the microenvironment surrounding the cell. Furthermore, cells without previous stem cell property were demonstrated to gain so when subjected to these specific set of extrinsic factors[42]. These data, taken together suggests that human colon cancer consists of a subset of cells with stem cell properties. These cells are most likely mutated normal intestinal cells with activated Wnt signalling pathway, express certain protein markers (like CD133, CD166, EpCAM, CD44) and their stem cell characters depend on certain external factors or surrounding microenvironment. However, there still remains substantial doubt as to which marker or combinations of which specific markers form the exact signature for these stem cells.

1.1.2. Morphogenetic pathways related to colon cancer:

The investigation of the molecular pathways regulating intestinal morphogenesis was hugely promoted by the studies related to the genetic background of familial syndromes, associated with a high risk of colorectal cancer like Juvenile polyposis syndrome (JPS), Familial adenomatous polyposis (FAP) and Cowden's syndrome. Approximately, 80% cases of FAP, present with mutations in *APC* (*adenomatous polyposis coli*) gene. Mutations in the genes coding for SMAD-4 and Bone Morphogenetic Protein (BMP)

receptor 1A are present in about 15-20% and 25-40% patients of JPS respectively. Cowdens disease causes hamartomatous neoplasm of skin and mucosa along with colonic polyps. PTEN (Phosphatase and tensin homologue) mutations are found in 80% such patients[2]. The sequence of events occurring in between normal colon to its progression to colon cancer is schematically shown in Figure1 (please see above).

1.1.2.1. Wnt pathway and APC:

Wnt glycoproteins (Wnt-1) are poorly soluble palmitoylated proteins mostly expressed towards the bottom of the crypts of Lieuberkuhn. These proteins regulate the Wnt signalling pathway (Figure 2) implicated in colorectal carcinogenesis[43]. The *APC* gene, present on the chromosome *5q21*, encodes protein APC. *APC* mutation is associated with 63% cases of adenomas and 60% cases of colorectal carcinomas[44]. In a study of 53 familial adenomatous polyposis syndrome (FAP) patients, it was found that >80% tumours had one mutation and in more than 60% patients there were at least 2 mutations in *APC*[45]. Most of the *APC* mutations occur at the mutation cluster region (mcr), starting at codon 1300[46].

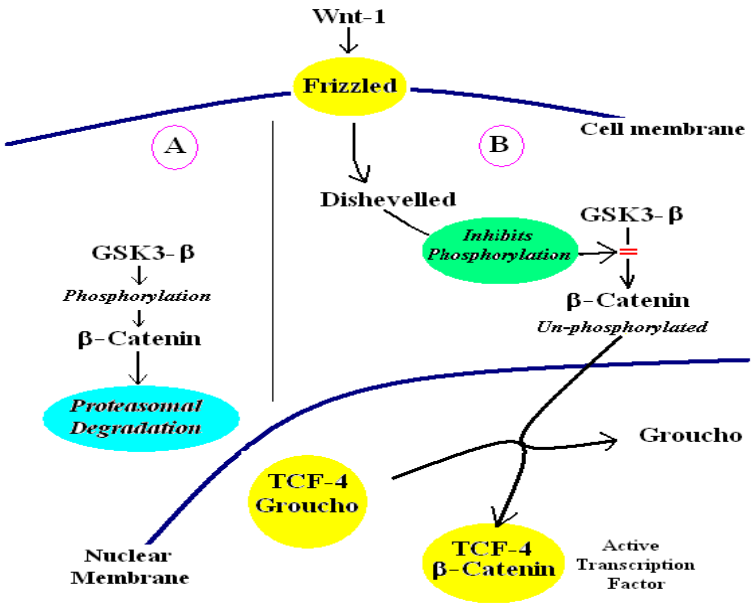


Figure 2: This figure demonstrates the canonical Wnt / β-catenin signalling pathway. Under normal circumstances, (Figure 2A) GSK3-β phosphorylates β-catenin, inducing the later to undergo proteasomal degradation. However, if this phosphorylation is

inhibited (Figure 2B) by Wnt signalling pathway (via activated frizzled and dishevelled respectively), β-catenin translocates in the nucleus, replaces protein Groucho (from Groucho-TCF4 complex) to form a complex with TCF-4. This molecular complex is an active transcription factor and is responsible for inducing several proteins related to cell proliferation.*(For further details please refer to Voutsadakis I. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2 [3].)*

A relatively large protein, APC harbours several functional domains. One of its important such domains consists of ten amino acid repeats, (three 15-amino acid repeats followed by seven 20-amino acid repeats) that bind β-catenin. Three Serine-Alanine-Methionine-Proline (SAMP) motifs are interspersed within the 20-amino acid repeats. These SAMP motifs are essential for the ability of APC to bind to Axin molecule. These functional domains play an important role in stabilizing APC as a part of a six protein complex (β-Catenin destruction complex) along with GSK3-β (Glycogen synthase kinase 3 beta), Axin, β-Catenin, Conductin and CK-I (Caseine Kinase I)[47-50]. The middle domain of APC is phosphorylated by GSK-3β. This phosphorylated segment and the SAMP motifs in the APC molecule have been found to be necessary for recruitment of β-catenin to the destruction complex[49,51]. Along with forming the β-Catenin destruction complex, APC also plays an active role in nuclear shuttling and sequestration of β-Catenin[52,53]. GSK3-β under normal circumstances, phosphorylates β-Catenin, both directly and indirectly (through activating CK). Phosphorylated β-Catenin undergoes ubiquitinization and proteasomal degradation by the UP (Ubiquitin-Proteasomal) System. A potent inducer of proliferation, Wnt-1 with co-receptor LRP-5 (Low density lipoprotein receptor related protein) ligates and activates cell surface receptor “Frizzled”, forming a complex, which in-turn activates protein “Dishevelled”. Activated “Dishevelled” along with GBP (GSK3-β binding protein) inhibits GSK3-β from phosphorylating β-Catenin. GSK3-β is also inhibited by other kinases like akt kinase (activated by K-ras), Protein Kinase-Cβ(PKCβ) and Integrin linked kinase (ILK) [54-56]. This prolongs the life of β-Catenin un-phosphorylated form. β-Catenin phosphorylation is also inhibited if the gene encoding for

β-Catenin (S37A mutant or serine at position 37 is replaced by alanine) or protein APC, undergoes mutation. Un-phosphorylated β-Catenin translocates into the nucleus and forms a complex with TCF-4 (T-cell factor) by replacement of the inhibitory protein "Groucho". β-Catenin-TCF-4 complex is an active transcription factor for several proteins related to cell cycle, proliferation, apoptosis and metastasis, like Cyclin D1, c-myc, CD44, IL-8 and Akt-1[3]. Among the genes, which are down regulated by this complex, are those coding for ZO-1 and Ephrin B[57,58]. Although the canonical view of APC/β-Catenin suggests that loss of function of APC results in nuclear translocation of β-Catenin, evidence suggests that specific post-translational modifications of β-Catenin molecule (example; tyrosine phosphorylation by activated c-kit) are necessary for its nuclear localization in response to Wnt stimulation[59,60].

β-Catenin induces transcription of genes (like those coding for ARF or alternative reading frame, which acts as a positive regulator of p53) in a TCF4 independent manner, possibly with the help of other transcription factors like Fox O (Forkhead Box O)[61]. β-Catenin acts as a bridge between E-Cadherin and the Actin cytoskeleton and thus maintains the integrity of the adherens-junction, the main inter-cellular adhesion point[62]. The cytoplasmic tail of E-cadherin becomes structured only when bound to β-catenin. This association begins in the endoplasmic reticulum and is found to be essential for the effective transport of cadherin to the cell surface for the formation of adherens junction (AJ, cell – cell adhesion complex, maintaining tissue homeostasis) complex. At the plasma membrane, this cadherin–β-catenin complex recruits α-catenin. α-catenin physically links the cadherin – catenin complex to actin molecules and this is essential for AJ formation and function[63]. β-Catenin also inhibits NF-κβ (promotes proliferation and angiogenesis) by direct interaction[64]. These actions of β-Catenin are anti-carcinogenic by inhibition of metastases, proliferation and angiogenesis and stabilisation of inter-cellular interactions. APC can also promote cell apoptosis in a Caspase-8 dependent manner, which is independent of β-Catenin. This is impeded in

truncating mutations of the APC gene and can contribute towards colon cancer pathogenesis[65]. By using Dickkopf1 (Dkk1), a specific inhibitor of the Wnt-LRP5-Frizzled complex, Pinto and colleagues[66] reported that Wnt signalling pathway is also essential in adult normal intestinal epithelium for the formation of normal crypt villus units by maintaining a correct balance of cell positioning, proliferation and differentiation. The Wnt pathway interacts with the Notch pathway and regulates proliferation and differentiation of stem/progenitor cells in the crypts[67].

1.1.2.2. K-ras and related pathways:

Activating mutation of *K-RAS* oncogene is found in about 50 % cases of colorectal cancer[68,69]. K-ras is involved in 3 important pathways related to colorectal cancer (as shown in Figure 3, below):

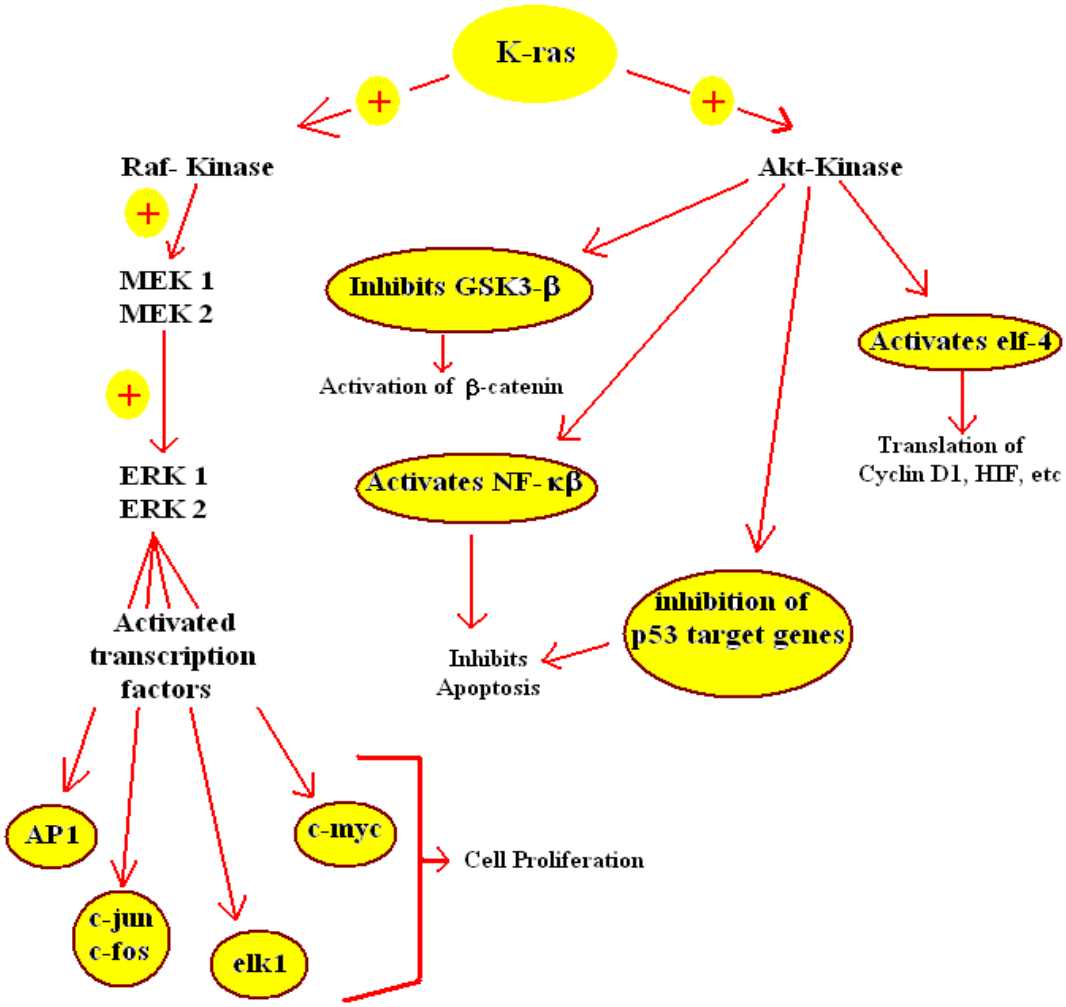


Figure 3: This figure demonstrates the mechanisms by which K-ras activates several transcription factors like AP1, elk1, c-jun – c-fos complex and c-myc to promote cell proliferation through activation of Raf-Kinase and inhibits apoptosis, prevents proteasomal degradation of β-catenin and induces proteins related to cell proliferation through activation of akt-kinase. *(For further details please refer to Voutsadakis I. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2* [3]*.)*

- Raf/MAPK kinase pathway
- PI3-K/akt kinase pathway
- Ral nucleotide exchange factor and Phospholipase Cε.

i) Raf / MAPK kinase pathway

In this pathway, K-ras activated Raf kinase (a form of Mitogen Activated Protein Kinase or MAPK) activates a cascade of kinases involving MEK-1 and 2, ERK 1 and 2 (extracellular signal regulated kinases 1 and 2), JNK 1 and 2 (c-jun N-terminal kinases 1 and 2) and finally activating transcription factors like AP-1, complex of c-jun and c-fos, elk-1 and c-myc. Of note, AP-1 and the complex of c-jun and c-fos promote cell proliferation[70,71]. MAPK activation by K-ras is also associated with up regulation of p-glycoprotein, which plays active role in imparting multi-drug resistance to colon cancer cells[72]. This pathway is also responsible for blocking TGF-β induced cell cycle arrest by inducing proteasomal degradation of p27, a cyclin dependent kinase (CDK) inhibitor[73]. In a review article, published by our group, on the role of 15-lipoxygenase-1 (15-LOX-1) in colon cancer, we have discussed that this may be a link between fatty acid metabolism pathway and colon cancer and may explain a pro-carcinogenic effect of 15-LOX-1[74].

ii) PI3-K/akt kinase pathway

In this second pathway, PI3-Kinase (a serine threonine kinase) is activated by K-ras, which in turn activates the kinase akt or PKB (Protein Kinase B). Downstream substrates of akt kinase undergoing inhibitory phosphorylation are GSK-3β, bad (pro-apoptotic

family member of bcl-2), caspase 9 (caspase in mitochondria dependent pathway) and transcription factors like FKHR and Fox O. This result into activation of β- Catenin (which may have anti-carcinogenic roles by inhibiting NF-κβ promoting cell cycle arrest by activating Fox-O and yet may be pro-carcinogenic by formation of a complex with TCF-4), inhibition of apoptosis and inhibition of cell cycle arrest respectively. Kinase IKK, another downstream substrate of kinase akt undergoes activation following phosphorylation, which results into subsequent phosphorylation and proteasomal degradation of I-κβ. (Inhibitor of NF-κβ). As NF-κβ is a potent pro-survival factor and inhibitor of apoptosis, this mechanism of inhibiting I-κβ is pro-carcinogenic[75,76]. Kinase akt also phosphorylates mdm2, resulting to its translocation into the nucleus, which in turn inhibits transcription of p53 target genes. Downstream substrates of kinase akt, which are activated, are mTOR (mammalian target of rapamycin) and PTEN (Phosphatase and Tensin at chromosome 10). Activated mTOR maintains elF4 in activated state by inactivating 4E-BP. This leads to translation of proteins like cyclin D1 and Heat inducible factor (HIF), which have been linked to proliferation of cells and colorectal carcinogenesis. PTEN is inhibitory to the PI3-K/akt pathway and therefore on activation completes a negative feedback loop[77-79].

Thus, interestingly both the Raf/MAPK and the PI3-K/akt pathways have pro and anti carcinogenic roles in colorectal cancer[3].

iii) Ral-nucleotide exchange factor and Phospholipase Cε (PLCε)

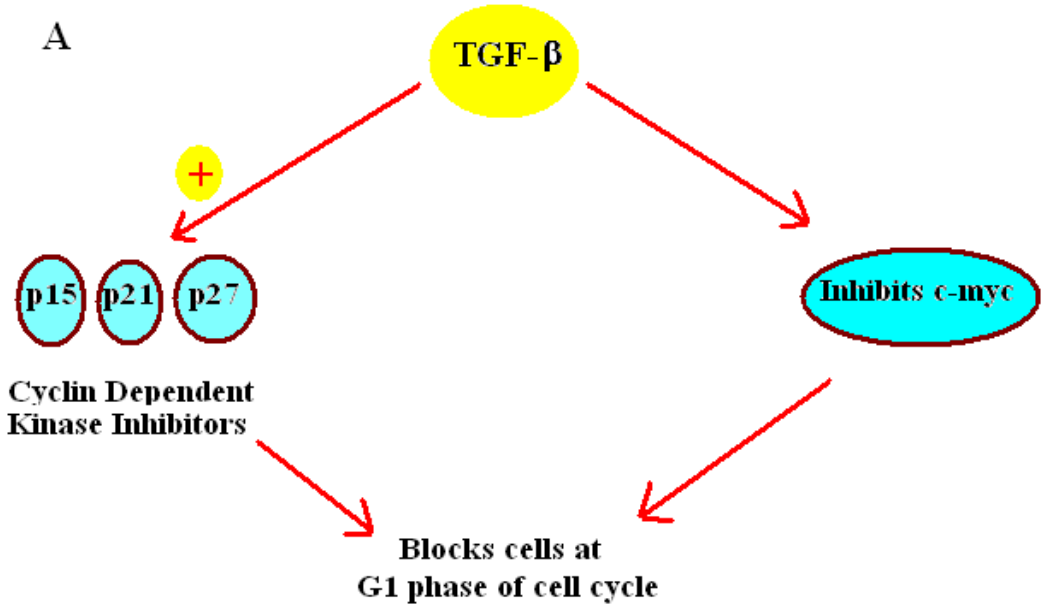
Ral nucleotide exchange factor activation by K-ras promotes cytoskeleton re-modelling and cell proliferation. PLCε activates the receptors involved in calcium signalling[80].

1.1.2.3. Notch Pathway:

The notch receptors are trans-membrane proteins and concerned with positive and negative regulation of proliferation and differentiation respectively, in the intestinal stem/progenitor cells. Notch and Wnt signalling function together to co-ordinate bHLH transcription factor dependent cell fate specification[81]. There are 4 types of notch receptors in vertebrates, Notch 1 –4. The ligands for these receptors are Delta like (Dll 1, 3 and 4) and Jagged (1 and 2) family ligands. The reaction between ligand and receptor is influenced by sugar transferases (from Fringe gene family) – Lfng, Mfng and Rfng[82-84]. Ligand-receptor complex activates the Notch protein and Notch protein is cleaved by TACE/ Kuzbanian proteases and γ–secretase. This separates an intracellular segment of the Notch protein called $Notch^{IC}$, which translocates into the nucleus and binds to DNA binding protein RBPJκ/CBF-1 and to the transcriptional regulatory proteins to control transcription[85]. Math-1 (Hath-1 in humans) is a bHLH transcription factor in mice. It induces differentiation of enterocytes, goblet cells, entero-endocrine cells and Paneth cells[86]. Notch pathway interruption by deletion of RBPJκ/CBF-1or inhibition of γ–secretase causes a rise in Math-1 and premature conversion of crypt cells to goblet cells (i.e increased differentiation). Notch has been found to activate Hes1 (bHLH suppressor) which in-turn perturbs the transcription of Math-1 and two CDK inhibitors, p27 (Kip1) and p57 (Kip2). This decreases differentiation and increases proliferation respectively[67,87]. Colorectal cancer specimens from human patients have been found to express increased levels of jagged ligands, Notch-1, Lfng, Hes1, low levels of Hath-1 (human orthologue for mouse Math-1) and no goblet cells. This proves that the Notch system is activated in colorectal cancer and can be a possible target for therapeutics in cancer management[88,89,90].

1.1.2.4. TGF - β Pathway:

Transforming growth factor – beta (TGF-β) is a protein molecule and is a part of the TGF-β superfamily, which also includes other proteins like inhibins, activins, bone morphogenetic protein (BMP) and Vg-1. Three isoforms of TGF-β have been identified: TGF-β1, TGF-β2 and TGF-β3. TGF-β molecules are secreted as large protein precursors, consisting of an N-terminal signal peptide (20-30 amino acids, required for secretion from the cell), an intermediate latency associated peptide (LAP) and a C-terminal segment comprising of 112-114 amino acids. The later becomes the mature TGF-β after proteolytic dissociation from the LAP segment. In the mature form, TGF-β exists as a dimer. A mature TGF-β molecule has nine cysteine residues of which the first eight forms a "cysteine knot", a structure characteristic of the TGF-β superfamily. The ninth cysteine bonds with a similar cysteine residue from another TGF-β to secure the dimer. TGF-β has anti-proliferative properties on normal epithelial cells. They are secreted by a wide variety of normal cells, including cancer cells, where the production of this protein factor is greatly exaggerated[91-93].



B

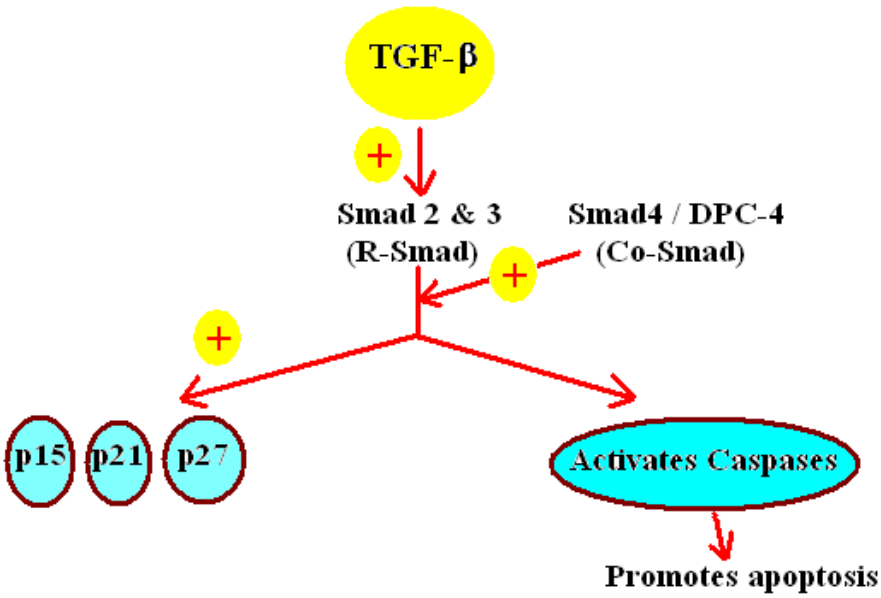


Figure 4A and 4B: This figure demonstrates the anti-carcinogenic roles of TGF-β in colon cancer by promoting cell cycle arrest at G1 phase (4A) and apoptosis (4B).*(For further details please refer to Voutsadakis I. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2* [3]*)*

TGF-β manifests its properties through cell surface receptors called TGF-β receptors (TβR). Two types of such receptors are found on vertebrate cells, TβRI and TβRII[94].

TGF-β blocks the cells at G1 phase of cell cycle by three vital mechanisms[95,96].

- Promoting production of p15 and p21, which blocks cyclin-CDK complex required for Retinoblastoma (Rb) protein phosphorylation.
- Stabilizing cyclin dependent kinase inhibitor (CDK-inhibitor), p27 by inhibiting its proteasomal degradation.
- Suppressing c-myc expression.

TGF-β induces apoptosis in cells by Smad pathway[97]. Smads are transcripion factors and are of the following types[97,98].

- R-Smad (receptor regulated Smad) or Smad 1, 2, 3, 5 and 9. Mutations of Smad 2 are found in about 6% cases of human colorectal cancer[99].

- Co-Smad (common mediator Smad) or Smad 4, also known as DPC-4 (Deleted in Pancreatic carcinoma-4). Mutations of Smad 4/ DPC-4 are found in 20% cases of colorectal cancer[100].
- I-Smad (inhibitor Smad) or Smad 6 and 7.

The above mechanisms are summarized in Figure 4. Binding of TGF-β to its cell surface receptors (TβRI and TβRII) initiates three main cascades of intracellular reactions. First, R-Smads (Smad 2 and 3) are activated. The activated Smad 2 and 3 along with Smad 4 promotes transcription of some putative target genes like *p15, p27, p21* (cyclin dependent kinase inhibitors) and *caspases* (apoptosis promoting enzymes) and represses transcription of *c-myc*. This mechanism is inhibited by TGIF (Smad Co-repressor / TGF-β inhibitory factor) and I-Smads (Smad 6 and 7). Second, activated TβRI and II promotes proteasomal inhibition[3]. Third, the K-ras and its related pathways are activated directly or indirectly, the later via activation of PI3-K/akt kinase pathway[101]. Thus, TGF-β demonstrates both pro and anti-neoplastic effects in colon cancer, with the cell fate being decided by the ratio of Smad 3 to akt-kinase[102].

Interactions between pathways related to TGF-β and K-ras

- K-ras activated MAPK/ERK pathway promotes proteasomal degradation of SMAD4, which inhibits TGF-β pathway[103].
- K-ras stabilises TGIF (TGF-β inhibitor) by inhibiting its degradation[104].
- TGF-β itself activates *K-RAS* directly or indirectly (via PI3-K/AKT-kinase pathway activation) as discussed above.

Thus, with intact Smad signalling, anti-carcinogenic effects predominate. However, debilitating mutations in Smads shift the balance towards K-ras pathway and are therefore pro-neoplastic in colon cancer.

Interactions between Wnt/β-Catenin pathway and TGF-β

- Axin, a component of β-Catenin/APC/GSK-3β 6 protein complex activates Smad 3 by direct interaction[105].
- Some complex co-operation exists between transcription regulation of target genes by Smads and β-Catenin/TCF-4 complex[106].

1.1.2.5. Bone Morphogenetic Protein (BMP):

These protein growth factors are related to the TGF-β (Transforming growth factor beta)family and are concerned with embryonic development and different cellular functions. Till date, nearly 20 BMP family members have been identified[107]. BMP activity was first described in 1965[108]. Purification and sequencing of Bovine BMP-3 (osteogenin) and cloning of human BMP-2 and 4 were reported subsequently[109-111]. Apart from playing a central role in formation of bone and cartilage, BMPs are also implicated in several non-osteogenic developmental processes like epidermal induction, development of neural crest and induction of sympathetic adrenergic phenotype[112-114]. BMP 2, 4 and 7 have been found to be absolutely essential for life through gene-knockout experiments in mouse models[107].

BMP signals through serine-threonine kinase receptors, which are composed of I and II subtypes. Three type I subtypes (IA BMP/ALK-3, IB BMP/ALK-6, IA Activin/ALK-2) and three type II subtypes (II BMP, II Activin and IIB Activin) have been discovered so far[115-120]. Of these receptors, type IA, IB and II BMP receptors are specific for the BMP ligands. After ligand binds to the receptor, a heterotetrameric receptor complex forms (consisting of two type I and two type II receptors)[121]. This receptor complex phosphorylates and activates Smad 1, 5 and 8. These phosphorylated Smad proteins separate from the receptor and after forming a complex with Smad 4 in the cytoplasm, translocates into the nucleus and act as a transcription factor[107,122-124]. Noggin and some cystine containing molecules bind to BMP 2, 4 and 7 and block BMP signalling[125]. Smad

6 acts as a negative regulator for the BMP signalling pathway by binding to type I BMP receptors, inhibiting phosphorylation of the Smad 1 and 5 and formation of complex with Smad 4[126]. Tob, a member of the anti-proliferative family of proteins, binds to BMP activated Smad 1 and 5 and blocks the BMP signalling pathway[127].

BMPs have been associated with the pathogenesis of some solid tumours. Up-regulation of BMP receptors (IA and II) and BMP-2 mRNA levels along with bi-allelic loss of *Smad 4* (also called *DPC-4* or *deleted in pancreatic carcinoma-4*) were associated with pancreatic cancer.

The protein, BMP-2 was also found to enhance proliferation of pancreatic cells[128-131]. Hardwick *et al* reported that BMP-2 has anti proliferative effects on normal colonic epithelial cells (by promoting apoptosis and differentiation) and found that BMP-2, pSmad1, Smad 4, BMP receptors IA, IB and II were predominantly expressed at the epithelial surface of the normal adult colonic tissue[132]. The issue of BMP in colorectal cancer becomes relevant when mutations in genes coding for Smad 4 and BMP receptor IA were found to be associated with Juvenile polyposis syndrome, which is an autosomal dominant condition presenting with multiple gastrointestinal hamatomatous polyps, with a risk of developing colon cancer[133-135](about 12 fold as compared to normal).

Enhanced formation of new crypts and lesions in the intestine, similar to the hamartomatous polyps in JPS, resulted from ectopic expression of Noggin[136] (BMP signalling inhibitor). Beck and colleagues used colon cancer cell lines (SW480 and HCT116) and human colon cancer specimens to confirm the role of BMP signalling in colon cancer and found that the signalling pathway is mostly intact and moderately growth suppresive in human colon cancer specimens and cell lines[137]. The Wnt pathway through activation of β-Catenin acts as a positive regulator and the BMP signalling acts as a negative regulator of the intestinal stem cell proliferation[57,138-140]. In a recent scientific paper Lombardo and colleagues have reported that BMP4 (bone morphogenetic protein 4) promotes differentiation and apoptosis in a colon cancer stem

cells in a selective group of tumours where simultaneous mutation in *Smad4* (mouse model) or constitutive activation of PI3-K/ akt kinase pathway is absent. They have also reported that BMP4 sensitises these tumours to chemotherapy with 5-fluorouracil and oxaliplatin in mice[141].

1.1.2.6. Pathways related to p53:

The transcription factor protein, p53 binds to DNA segments, which are comprised of specific base sequence, RRRC(A/T)(T/A)GYYY, where Y is a pyrimidine and R represents a purine. The carboxy-terminal of p53 has 3 lysine residues, which interacts non-specifically with several DNA segments. Acetylation of these lysine residues by p300 prevents such non-specific DNA binding and possibly regulates p53 interaction with target genes. Poly-ubiquitinization of mono-ubiquitinated p53 (by mdm-2) is also promoted by p300, which results ultimately into proteasomal degradation of p53.

Of the many important genes, transcriptions of following genes related to colon cancer are influenced by p53:

- Inhibition of transcription of anti-apoptotic *bcl-2*, *bcl-xn* and *survivin.*

Activation of transcription of *Siah-1.* Siah-1 ubiquitinates β-Catenin leading to its proteasomal degradation and therefore prevents formation of β-Catenin -TCF-4 complex (functions of this has been discussed in the Wnt pathway and *APC*)[142,143].

Thus p53 mainly results in increased apoptosis and cell cycle arrest. In nearly 50% cases of colon cancer, some form of loss of function mutation of *p53* is identified, majority (95%) of which affects the DNA binding site of this protein[144].

Regulation of *p53* is an interesting topic in cancer research. The regulation may be:

1. Negative: Inhibition of p53 is caused by the following E3 ubiquitin ligases which ubiquitinates p53 leading to its proteasomal degradation:

- Mdm-2 (mouse double minute 2, hdm-2 for humans)[145].
- COP1[146], Pirh2[147], ARF-BP1/Mule[148].

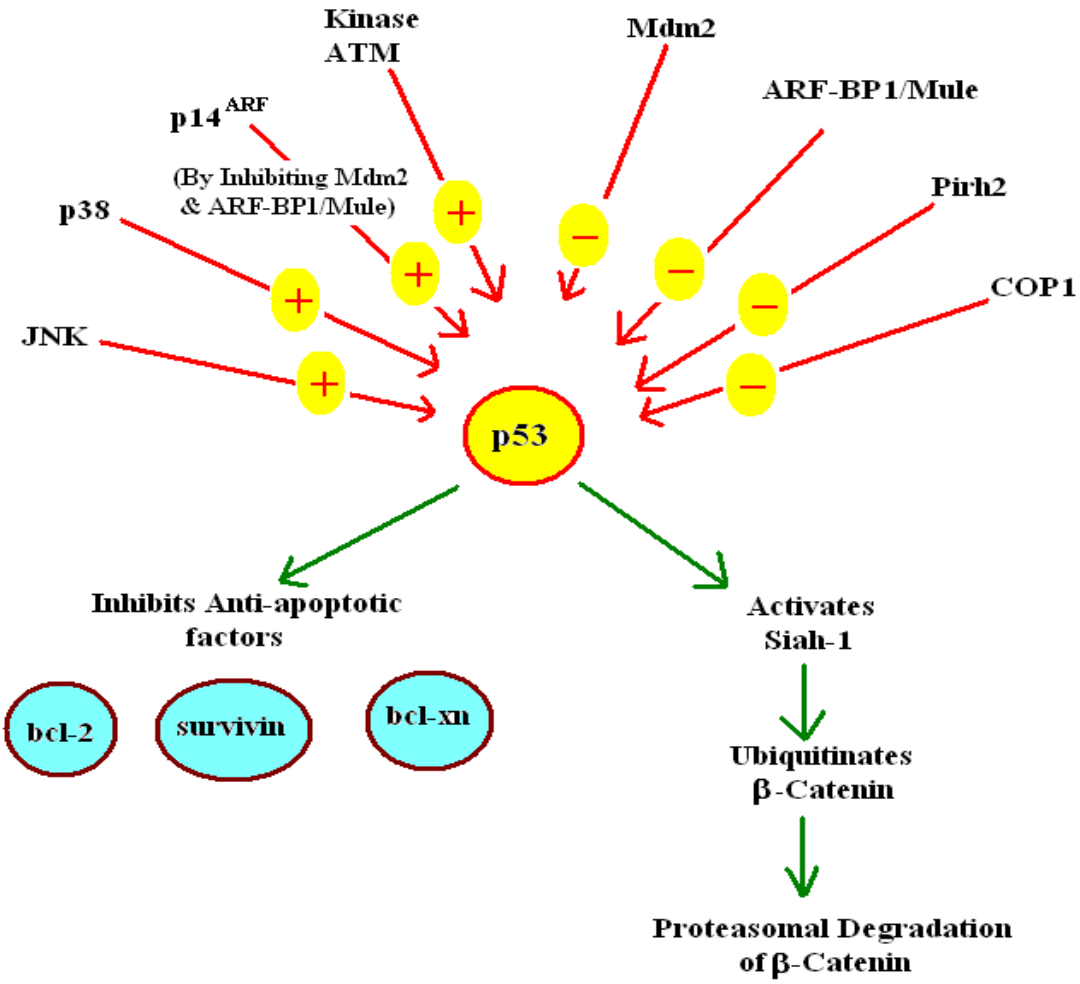


Figure 5: This figure demonstrates mechanisms related to the regulation of p53 and its anti carcinogenic role by inhibiting anti- apoptotic factors (bcl-2, surviving and bcl-xn) and promoting proteasomal degradation of β-catenin.*(For further details please refer to Voutsadakis I. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2* [3]*.)*

2. Positive: Following may cause positive regulation:

- $p14^{ARF}$, which inhibits both mdm-2 (hdm-2) and ARF-BP1/Mule.
- Kinase ATM (Ataxia telangiectasia mutated) both directly and indirectly phosphorylates and activates p53, the later by activation of chk1/2 kinases.
- GSK3-β, p38 and JNK[145].

The above mechanisms are summarized in Figure 5.

1.1.2.7. Colon cancer and its relationship with fatty acid metabolism:

Polyunsaturated fatty acids like arachidonic and linoleic acids (n-6 fatty acids) are pro-carcinogenic unlike fish oils or n-3 polyunsaturated fatty acids, which may be protective[1-3]. Oxidative metabolism of arachidonic and linoleic acids by the lipoxygenase (LOX) and cycloxygenase (COX) enzymes generates metabolites, shown to contribute to the pathogenesis and progression of various cancers, notably colorectal[149]. Moreover, therapeutic targeting of n-6 fatty acid metabolism by agents, such as non-steroidal anti-inflammatory drugs (NSAIDs), can inhibit tumour formation. Extensive research is on-going in this area and several distinct new therapeutic targets are emerging in the 15-lipoxygenase pathway for cancer chemo prevention in man[150]. Here we describe the role of the15-lipoxygenase-1 (15-LOX-1) pathway of n-6 fatty acid metabolism in the patho-etiology and progression of colorectal cancer and highlight recent progress in developing novel therapeutics based on its manipulation.

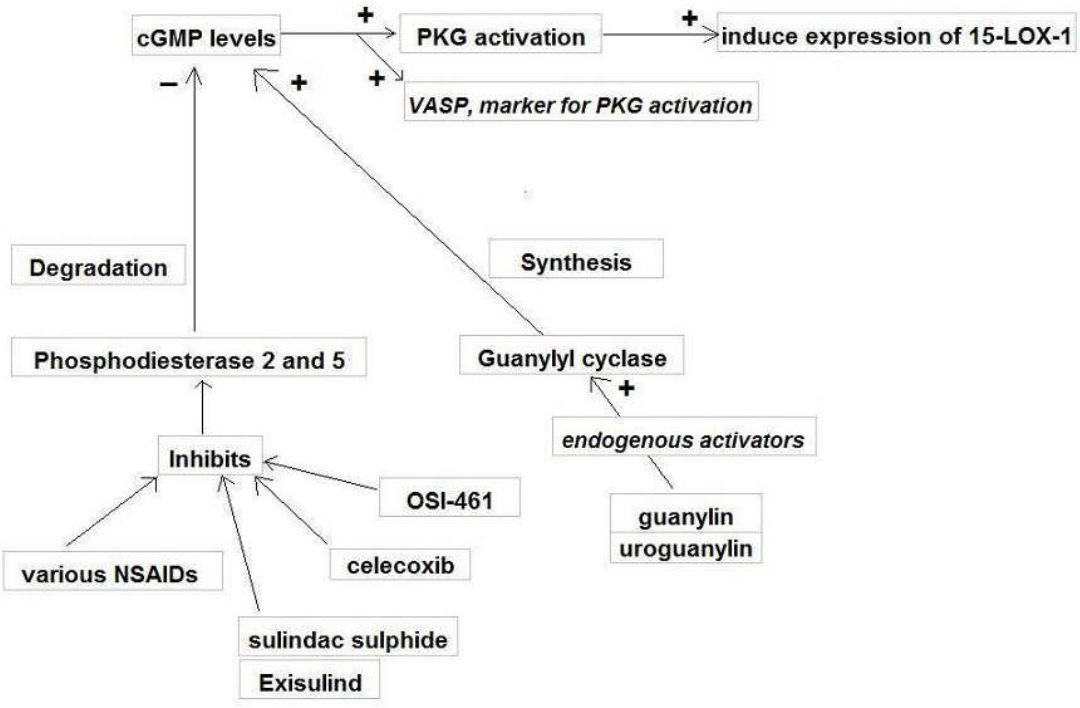


Figure 6A: Mechanisms related to Non-steroidal anti-inflammatory drugs mediating 15-lipoxygenase-1 induced apoptosis are demonstrated in this figure. (For further details please refer to Bhattacharya S et al. 15-Lipoxygenase-1 in Colorectal Cancer: A Review[74].)

Arachidonic acid metabolism by lipoxygenase in platelets was demonstrated in 1974[151]. Rabbit reticulocyte lipoxygenase was described for the first time in 1975[152]. The lipoxygenases are cytosolic dioxygenases, composed of a single polypeptide chain (mol wt. 75-80 kDa) and react with polyunsaturated fatty acids to produce the corresponding hydroperoxyderivatives. They have 2 domains, the smaller N-terminal beta barrel domain and the larger C-terminal catalytic domain. A single non-heme iron is liganded to the later domain by four histidines and the C-terminal isoleucine. In the active state of this enzyme, the iron is in ferric form and reduced to ferrous form in the inactive state. The beta barrel at the N-terminal has homology to a similar domain at the C-terminal of the mammalian lipase, which is involved in substrate acquisition. This suggests a similar role of this domain in the lipoxygenase[153].

Although several inter-conflicting reports has been published, convincing recent evidences from several *in vitro* and *in vivo* studies, especially the ones in the latter category from Nixon *et al*[154], Shureiqi *et al*[155] and Wu *et al*[156] makes us believe that the overall effect of 15-LOX-1 and its 13-S-HODE metabolite is largely anti-carcinogenic in colorectal cancer. Several of these reports also strengthen the view that up-regulation of 15-LOX-1 by specific targeted gene therapy has the potential to induce apoptosis and decrease proliferation in colon cancer cells, which makes this molecule a potential therapeutic target. The anti-carcinogenic role of 15-LOX-1 gains further support from the reports supporting its inhibitory effects in angiogenesis, metastasis and facilitatory effects on apoptosis and cell cycle arrest by increasing levels of p53.

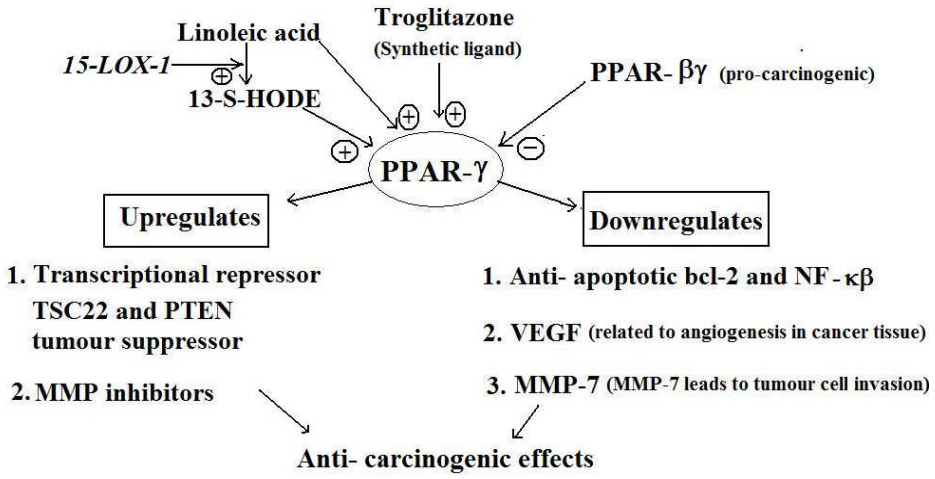


Figure 6B: This figure demonstrates the effects of PPAR and 15-LOX-1 in colon cancer.

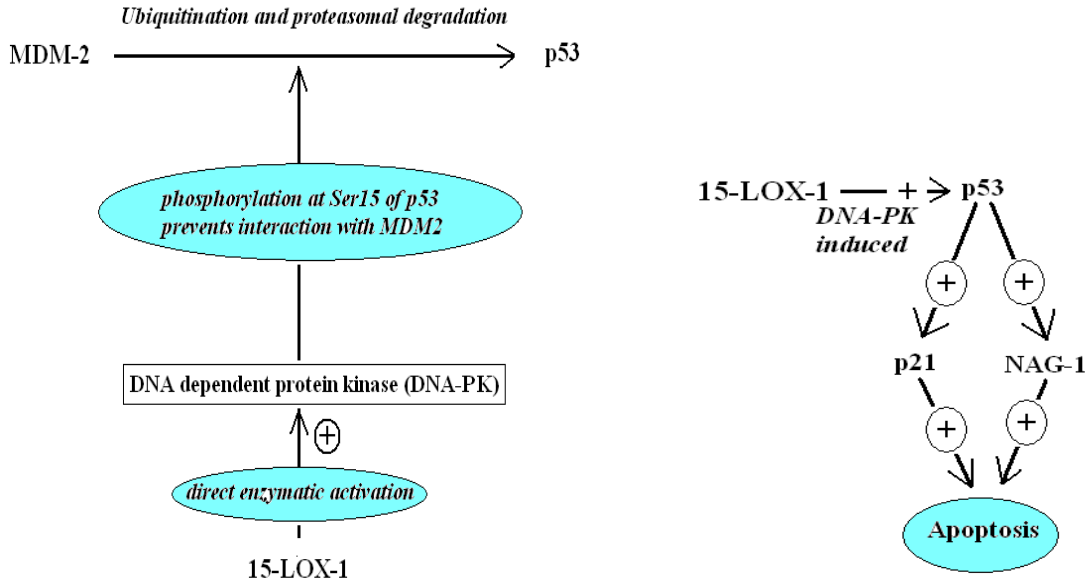


Figure 6C: This figure shows 15-lipoxygenase induced apoptosis dependent on p53, p21 and NAG-1.(For further details please refer to Bhattacharya S et al. 15-Lipoxygenase-1 in Colorectal Cancer: A Review [74].)

However *in vitro* reports of suppression of p21 and inhibition of PPAR-γ by 13-S-HODE by Yoshinaga *et al*[157] and Hsi *et al*[158] respectively might be contradictory to these findings. Of note, both of these evidences were based upon *in vitro* studies and results from animal experiments, confirming the relationship between 13-S-HODE with p21 and MAPK are still lacking. Interestingly, *in vivo* studies by Shureiqi *et al*[155] in nude mouse

xenograft model has demonstrated that 13-S-HODE exerts inhibitory effects on PPAR-δβ, an effect which can explain the anti-carcinogenic effects of 13-S-HODE on PPAR-γ and NSAID induced apoptosis in colorectal cancer cells. A detailed review on this subject has been published by our group and are summarised in the Figures 6 A, B and C[74].

1.1.3. Biomarkers and colon cancer:

The term "biomarker" in cancer research refers to substances or processes indicating tumour at primary or secondary (metastatic) sites[159,160]. This could either be molecules secreted by cancer cells, also known as "cancer cell secretome" or a specific response of the body to the presence of cancer[160,161]. The objectives behind cancer biomarker research is therefore to invent simple, cost effective, non invasive or minimally invasive investigations, preferably using easily accessible body fluids (blood, urine or saliva), which would indicate the risk of cancer in asymptomatic healthy individuals, allow detection of cancer at its early stages, provides us with simple tools to monitor disease progression and guide us in its management, along with giving us information regarding its prognosis[162-164]. Molecular biomarkers are classified under several broad classes according to their role[165,166].

1. Diagnostic markers, used for differentiating between cancer and non-cancer cases.
2. Prognostic markers, predicting disease outcome.
3. Markers used to monitor therapeutic response, or response to certain treatment (example chemotherapy).
4. Predictive markers, used to guide the clinician to choose the best treatment modality, or the therapeutic intervention to which the cancer is most likely to respond.
5. Screening marker, providing important information regarding disease risk in general population aiming towards early detection of cancer[160,167].

Biomarkers can be further classified into,

1. Cancer specific markers, presence of which indicates the likelihood of presence of a specific cancer. (example CEA, CA19-9, CA125, etc)[168].
2. Tissue specific markers, which are secreted by a specific tissue (example Prostate specific antigen, secreted by prostrate, not only raised in prostate cancer, but also in other prostate related conditions and even with digital rectal examination). However, its serial measurement is a very useful tool to identify response of prostate cancer to treatment[168,169].

Although, extensive literature exists on potential biomarkers, only a handful of them have clinical application in diagnosing and monitoring cancers. Furthermore, several shortcomings exist with the currently available biomarkers; for example, poor sensitivity and specificity limits their clinical importance[163]. Assays used to detect over expressed protein biomarkers from serum are techniques like immunoassays, with potential drawbacks like necessity of specific antibodies and decrease in sensitivity when molecules or cells of low abundance are detected in a medium (example, blood) where other proteins or cells are present in abundance. RT-PCR used to detect mRNAs, are based on the availability of special RNA isolation kits and RNA preservation techniques. Although these are highly improved, mRNA instability poses a major problem. Furthermore, carry-over contamination, detection of pseudogene-expressed mRNA, illegitimate transcription and marker expression by normal cells affect the specificity of this technology[169-171].

Along with facing obstacles like relatively low abundance of target proteins in biological fluids, association of target proteins with carrier proteins and labile nature of these proteins, there are other biological variabilities and those associated with collecting specimens and their preparation[172,173]. Although most easily accessible, plasma or serum proteomics becomes challenging due to extreme complexity and dynamic changes in structure and function of these proteins in response to physiological and pathological changes. Blood contains numerous proteins, essential for maintaining body

homeostasis[159,170]. Proteins from “cancer secretome” are also added to this pool. Cleavage by proteases, alternative splicing of mRNA, single nucleotide polymorphisms (SNPs), post translational modifications and other chromosomal genetic variations result in many of these proteins existing in multiple forms. Complex disease processes like cancers are likely to further increase this number of protein molecules due to mutations and alterations in splicing and abnormal post translational mechanisms [174]. Microarray based exon profiling is used for studying alternative splicing. Post translational modifications like acylation, phosphorylation, glycosylation, oxidation and cleavage are studied by multiplexed panels of biomarkers or proteomic techniques based on mass spectrometry or protein microarrays[164,175,176]. As already mentioned above, relatively low abundance of potential biomarker proteins amidst other proteins in blood complicates the processes involved in identifying and measuring them. Concentrations of known plasma proteins range from 30 – 50 mg / ml (example, albumin). Less abundant proteins have lower concentrations (1 ng/ml for some cytokines). Proteins related to cancer secretome are expected to be of even lower concentration[159,169]. Modern advancements in protein microarray and mass spectrometry (MS) have given us the capacity to detect proteins at sub-nanogram level[172]. Conventional centrifugation, centrifugal ultra-filtration, chromatographic techniques, utilizing nanoporous substrates, different immune-affinity and precipitation methods are often used to deplete more abundant “unwanted” proteins and for enriching low concentration biomarker “target” proteins[177-180]. An important drawback in these approaches is the associated inadvertent loss of several low abundance proteins or peptides, complexed to carrier proteins, such as albumin[181]. Attempts to isolate these sequestered proteins by affinity harvesting of albumin from blood samples and more recently developed, a “nano-particle technology”, revealed that they could offer significant information[182-186]. The later used hydrogel nano-particles as affinity bait to capture target proteins. Thus by using nano-porous core-shell enabled size fractionation; proteins are protected against actions of endogenous and exogenous proteinases. The main advantage in this technique is the ability to sequester target

proteins (on affinity bait, in the centre of nano-particle) away from the carrier proteins like albumin, thus enabling concentration of analytes[174]. In this technique, blood can be used directly in vacutainers, containing nano-particle solutions, which is likely to solve the problems related to short half life and lability of proteins[187]. All biological samples are affected by variabilities related to sex, age, ethnicity, diet, exercise and physiological states (for example; hormonal imbalance and fasting) along with changes imparted by diseases[188]. These factors could be resolved by referencing databases from systematic proteomic analyses of body fluids by projects such as HUPO (Human Proteome Organization) Plasma Proteome Project[174].

The data obtained from genomic and proteomic analyses are essential for understanding underlying factors involved in human diseases, especially complex processes involved in cancer biology. This data is also pivotal for the discovery of diagnostic biomarkers, as well as for developing further insight into the metabolic effects mediated by signalling molecules. “Omic” approaches, namely genomics, transcriptomics, proteomics, metabonomics etc are used for analyzing all classes of biological compounds, from genes through mRNA to proteins and metabolites, required to develop a broader view of the complex biological system, including the pathology of diseases. While the data obtained from genomics explain the disposition of diseases (i.e., presence or absence of genes related to increased risk of acquiring a certain disease), several other mechanisms that are not gene mediated may be involved in the onset of disease as well. A single gene can be processed to result in several different mRNAs and proteins, which directly determine different cellular functions. Variations in metabolic effects and metabolic fluxes as the downstream result of changes in gene-expression and protein translation, are expected to be amplified relative to changes in the respective transcriptome and proteome. Interestingly, these changes are extremely time dependent. Thus measurements and determinations of metabolite content at a single time-point can be misleading as these fluxes vary quickly. Therefore, in order to understand the complete disease process, the entire set of “omic” techniques are

important. While genomics/transcriptomics enables assessments of all potential information regarding genes and m-RNAs, proteomics enables us to assess the programs that are actually executed by providing important information about the proteins produced and metabonomics will display the results of such executions[189]. Functional analysis of genes and their products in relation to health and disease, constitutes a powerful and novel approach since the expression levels of multiple genes and their resultant proteins can be analyzed simultaneously. Techniques related to functional genomics like DNA microarrays and classical proteomic approaches (extraction of different proteins from the tissue and identifying them separately) hold great promise for the study of complex biological systems, along with having on-going applications in molecular medicine. High throughput analysis of tissue is now possible using these techniques, as being complementary to each other they lead to a better understanding of the regulatory events involved both in physiological and disease processes. Proteins are excellent targets for being prognostic, diagnostic and therapeutic factors in diseases like cancers. Thus, proteomic approaches, such as mass spectrometry (MS), two dimensional gel electrophoresis (2-DE) and two dimensional liquid chromatography (2-DL), which allow simultaneous measurement of expression levels of hundreds of proteins and comparing them in different subsets, represent powerful tools for inventing novel therapeutic and prognostic targets and biomarkers and in-depth studying of cellular metabolism and protein expressions[190].

Screening techniques ideally applies to diseases with significant incidence, prevalence, morbidity and mortality along with a natural history which can be timely intercepted and influenced by medical or surgical interventions, thereby influencing the course of the disease. Colorectal cancer fulfils most of the above criteria. In most cases, it has a long precancerous stage through adenomas or polyps. In the present day, implemented screening techniques play an important role in early detection of these precursors of colon cancer, its primary prevention and reduction of cancer related mortality[191]. Yet, nearly 70 percent of colon cancers present at advanced stages of disease (UICC III/IV)

with five year disease free survival rates of 60 percent or even lower[192]. Therefore introduction of more efficient screening tools is imperative for curative interventions.

Highly organized protein networks, both intracellular and cell-surface, are responsible for determining their functionality. Thus the function of a cell is determined at following levels:

- Genomic level, genetic codes deciding expression of a protein, chromosomal abnormalities, mutations, etc.
- Transcriptional level, influencing protein synthesis at m-RNA level.
- Proteomic level and post translational modifications.
- Toponomic level, which determines the interaction between different proteins which is highly organized quantitatively, timely and spatially[193-195].

We will classify the biomarkers according to the levels of cellular function they arise from.

1.1.3.1. Genomic level biomarkers:

A key mechanistic component for initiation and progression of colon cancer is instability at genomic and epigenetic levels, increasing the rate of accumulation of such damaging alterations, which permits the adaptations characteristic of a malignant tumour. An important example in this field is the role of Microsatellite Instability (MSI) in the pathogenesis of colon cancer[196]. Microsatellites are defined as repeated sequences of DNA, the length of which are highly variable from person to person; each individual having microsatellites of a set length. These repeated sequences are common, and normally present. The appearance of abnormally long or short microsatellites in an individual's DNA due to mutations in DNA repair genes is referred to as MSI[197]. Microsatellites are also known as simple sequence repeats (SSRs). MSI is one of the most promising molecular markers investigated to date, especially in relation to biomarkers for colon cancer and also in relation to many other malignancies. It is

estimated to be the cause for 15% of colon cancers. The remaining 85% of colorectal cancers develop from the chromosomal instability (microsatellite-stable or MSS) pathways, characterized by mechanisms like aneuploidy, allelic losses, amplifications, and translocations. It has been demonstrated by several authors that tumours with high-frequency MSI are characterised by high lymphocyte infiltration, tend to be more proximally located in the colon, poorly differentiated, mucinous and retain their native diploid state. In contrast, microsatellite stable tumours tend to be aneuploid and have no site predilection[198,199]. A number of studies have investigated the relationship between MSI status and prognosis in colon cancer patients. There is considerable difference in opinion in between different reports. An extensive systematic review of the published data by Popat and colleagues used standard techniques and estimated the prognostic significance of this phenotype with relation to colorectal cancer. From their analyses, the authors reported that tumours with MSI have a significantly improved prognosis. However, these tumours were found to be slightly more resistant to chemotherapy by Fluoro-pyrimidines, the later being the mainstay in colon cancer chemotherapy in the present day. Although, these findings point towards a benefit in pre-selecting colon cancer patients with or without MSI before making a treatment plan with adjuvant chemotherapy in locally advanced stage III disease, the authors suggested the need for further prospective randomised trials to validate this finding[200]. A more recent meta-analysis specifically studied the role of MSI status in predicting the response to chemotherapy in colorectal cancer patients. Interestingly, the authors showed that although, MSI tumours had slightly better prognosis than MSS tumours, there was no difference in “relapse-free survival” in patients with high MSI, whether or not they received chemotherapy. In sharp contrast, MSS patients demonstrated a much better response to chemotherapy, suggesting that MSI positivity in colon cancer predicts chemo-resistance[201].

Genome wide association studies (GWAS) may be considered as a significant approach for high-throughput identification of common, low penetrance alleles modifying the risk

for complex disease processes like colon cancer. These minor, yet common variations detected in DNA sequences are referred to as single nucleotide polymorphisms or SNPs. A small proportion (3-4%) of colon cancer results from familial syndromes caused by rare mutations in genes of high penetrance. Important examples in this are APC (adenomatous polyposis coli) syndrome, resulting in multiple colonic polyps, and the genes related to DNA MMR (mismatch repair) that causes Lynch syndrome. The later is an autosomal dominant disease causing early onset colon cancer along with a range of extra colonic tumours, characterised by carcinogenic mutations in DNA MMR genes: *MLH1* (MutL homolog 1), *MSH2* (MutS homolog 2), less frequently in *MSH6* (MutS homolog 6) and *PMS2* (post meiotic segregation 2). Methylation of MLH1 can also result in its deactivation resulting in sporadic colon cancer, not related to Lynch syndrome. These methylated forms have been reported to be found in peripheral blood and lymphocytes, along with tumour tissue[202-204]. Ligtenberg and colleagues identified a relationship in between *MSH2* methylation and colon cancer pathogenesis. The authors reported that germline deletions at 3′-end of the epithelial cell adhesion molecule (EpCAM) gene, also known as tumor-associated calcium signal transducer 1 or *TACSTD1*, located immediately upstream of *MSH2* were the possible carcinogenic changes, which abolished transcriptional termination of *EpCAM* and causing transcriptional read-through into *MSH2* resulting in over-expression of fusion transcripts in epithelial tissues. The level of expression of these fusion transcripts correlated well with the level of *MSH2* methylation in these patients[205,206].The human genome project and extensive linkage analysis suggest that the high penetrance genes responsible for such syndromes have essentially been identified. Many other familial clusters of colon cancer are present, which are thought to arise from mutation in genes of low penetrance. Through the GWAS technology, simultaneous linkage analysis of thousands of SNPs has been made possible. Data from these linkage analysis studies thus help us associate SNPs with certain disease processes[207]. These data have allowed us to better understand the role of common genetic variants in the cause of cancer and other

disease processes. In spite of its existence for last several years, it was in late 2007 when GWAS studies associated the first common low penetrance susceptibility variant with the risk of colon cancer[208]. Since then, many other such variants have been linked to this disease, as has been extensively reviewed by Goel and Boland[209]. Studies by two independent research groups reported that one of these SNP variants 8q24 rs6983267 is present in nearly 50% of Europeans and almost 100% African population, thus being quite common. Although, homozygosis of the G allele for this SNP raises the risk of colon cancer in the population by only 1.5 times, it demonstrates relative copy number increase during the actual process of development of tumour. Furthermore, this region acts as a "transcriptional enhancer" for Wnt signalling pathway, a key mechanism involved in colon cancer pathogenesis as discussed above[210].

K-RAS is adhered to the inner layer of the cell membrane and is mutated early and most frequently in colon cancer, among all the RAS family genes (K, H and N)[211]. As discussed above, mutations in K-RAS plays a significant role in transformation of normal colon cells to adenoma (Figure 1). The largest international study (RASCAL) in this respect reports association of its mutation with increased incidences of recurrence and death in colorectal cancer[69]. Poor outcome to chemotherapy has been reported with mutated *K-RAS*[212]. Early and frequent mutation and its detection in stool samples from colon cancer patients support its potential role as a biomarker. An assay based on detection of mutated *K-RAS* has been developed but not yet been introduced for clinical application[213].

As discussed above, *APC* regulates cell proliferation and is known to be a tumour suppressor gene in colon cancer. Mutations inactivating this gene are present in up to 70% sporadic cases of colorectal cancer (e.g. Familial adenomatous polyposis coli syndrome). Though this mutation occurs in early stages of the disease, its use as a diagnostic, screening or surveillance tool has not met much success[214].

Mutations in genes related to DNA mismatch repair result into defective DNA repair, causing genetic microsatellite instability. Halling and colleagues found an association

between MSI and lower 5-year survival in colorectal cancer patients[215]. However its independent role as a biomarker has not yet been established.

P53 is a tumour suppressor gene is involved with many important functions like cell cycle control, inducing apoptosis and repairing DNA damage[216]. Mutations in this vital gene are associated with transition of colorectal adenoma to carcinoma (Figure 1), resulting in expression of the defective p53 protein in the nuclei of colonocytes, detectable by immunohistochemistry[3]. Although *p53* mutation has been shown to result in shorter 5 year survival and worse prognosis by some studies[217,218], conflicting results have been reported in the literature[219]. Therefore its role as a biomarker is still under question.

Hyper methylation of specific genes is a frequent event in colon cancer[220]. Restriction enzyme-based qPCR, used to detect methylated forms of SEPT9 in serum samples diagnosed colon cancer with a sensitivity of 52% and specificity of 95%[221].

1.1.3.2. Transcriptomics:

The term “transcriptome” describes the entire set of RNA molecules produced in one or a population of cells and includes mRNA, rRNA, tRNA, and other non-coding RNAs. The term can also be applied to the total set of transcripts in a given organism, or in a particular cell type. Unlike the genome, which is fixed for a given cell line (excepting mutating cell lines) the transcriptome varies with external environmental conditions. As it includes all mRNA transcripts, the transcriptome represents the genes that are being actively expressed at any given time, with the exception of mRNA degradation (such as transcriptional attenuation). The study of transcriptomics, also known as “gene expression profiling” thus examines the expression level of mRNAs in a cell population, often using high-throughput techniques based on DNA microarray techniques for interpreting the functional elements of the genome, and also for understanding development and disease. This technique enables us to catalogue all transcripts in a defined subset of cells, determining the transcriptional structure of genes, splicing

patterns and other post-transcriptional modifications[222]. Various technologies have so far been developed to study the transcriptome, including hybridization or sequence-based approaches. The former involves incubating fluorescently labelled cDNA with custom-made or commercial microarrays. Specially designed microarrays with probes spanning exon junctions are used to detect and quantify distinct spliced isoforms[223].

In contrast to microarrays, sequence-based techniques directly determine the sequence of cDNAs . Use of Sanger sequencing of cDNA or EST libraries was found to be relatively low throughput, expensive and generally not quantitative[224,225]. To overcome these above limitations, tag-based methods were developed, including serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and massively parallel signature sequencing (MPSS). Although these approaches were found to be high throughput and provide precise, 'digital' gene expression levels, few disadvantages were apparent. Most of these techniques were based on expensive Sanger sequencing technology, and a significant portion of the short tags could not be uniquely mapped to the reference genome. Moreover, only a portion of the transcript could be analysed and isoforms were generally indistinguishable from each other. These disadvantages pose potential limitations to the use of traditional sequencing technology in annotating the structure of transcriptomes[226-229]. Recent development of novel high-throughput DNA sequencing method, known as RNA-Seq (RNA sequencing) has provided a new platform for mapping and quantifying transcriptome. Due to clear advantages of this technique over existing approaches, it is expected to revolutionize the study of transcriptomics, as apparent from its previous applications to *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Arabidopsis thaliana*, mouse and human cells, including its application in transcriptome profiling of stem cells[230-235]. Although still under constant development, RNA-Seq offers several key advantages over other existing technologies. First, unlike hybridization-based approaches, RNA-Seq is not limited to detecting transcripts that correspond to existing genomic sequence[236]. This makes RNA-Seq particularly attractive for non-model organisms with genomic

sequences that are yet to be determined. In addition, RNA-Seq can also reveal sequence variations (for example, single nucleotide polymorphisms or SNPs) in the transcribed regions[235,237]. A second advantage of RNA-Seq relative to DNA microarrays is that RNA-Seq has very low, if any, background signal. As it does not have an upper limit for quantification, correlating with the number of sequences obtained, it has a large dynamic range of expression levels (very high sensitivity) over which transcripts can be detected. In contrast, for genes expressed either at very high or extremely low levels, DNA microarrays lack sensitivity and therefore have a much smaller dynamic range. The results from RNA-Seq are highly reproducible and thus reliable, for both technical and biological replicates. Finally, because there are no cloning steps, RNA-Seq requires less RNA sample[230,233,235].

In relation to human cancers, gene expression profiling has been successfully used in breast cancer research. Specific subtypes of this disease (for example, tumours expressing HER-2 or Human Epidermal growth factor Receptor - 2) are found to have distinctive gene expression profiles and thus represent diverse biological entities associated with clear differences in clinical outcome[238,239]. Gene expression profiles have also been studied for different subgroups of breast cancer (with difference in Estrogen receptor status or tumour metastases in lymph nodes, expressed as lymph node status) with different clinical outcomes and responses to therapy, with clear differences in between these subgroups with regards to prognosis[240]. In a case-control study by van't Veer and colleagues, from the Netherlands Cancer Institute, one such profile, consisting of 70 genes, was developed using archived frozen tissue from a cohort of 78 node-negative patients with breast cancer[241]. The authors demonstrated that these rapidly relapsing tumours, following primary therapy had gene expression profiles that were quite distinct from those who remained disease-free after treatment. The authors further validated their results by applying these gene expression profiles to a second set of 295 frozen tissue specimens, collected from a new cohort of patients (including 61 patients from the previous group), yielding very similar results[242]. This 70-

gene profile was thus found to predict outcome in breast cancer more efficiently than traditional clinical criteria. These preliminary results thus suggested that transcriptomics may be regarded as a very powerful tool for predicting outcome and may also be considered as a guide to management.

Aneuploidy is associated with advanced colorectal cancer, increased incidence of colonic dysplasia in ulcerative colitis and poor survival in colorectal cancer patients[243-246]. Specific aneuploidies (example, trisomies) exert direct influence on the expression of genes in affected chromosomes. These changes result in distinct levels of genetic and protein expression, which correlate with the subsequent stages of colon cancer progression[247]. Simultaneous analysis of thousands of genes in tumour genomes is possible by microarray based gene expression profiling[248]. A search for possible biomarkers for colorectal cancer has been performed by several researchers by this technique[249].

Expression profiling has also been used to investigate the role of miRNAs (micro RNAs) in diagnosis of solid tumours, based on the expression signature of individual miRNAs. Robust evidence shows that along with differentiating in between normal and cancer tissue, expression profile of miRNAs can also distinguish between types and grades of different individual tumours. A systematic analyses by Lu *et al* demonstrated successful classification and diagnosis of different solid tumours based on specific mRNA expression pattern with 70% accuracy[250]. Silencing of 32 miRNAs along with over expression of 18 different miRNAs, with nearly 10 times over expression of 6 miRNAs was reported in colon cancer as compared to normal colon tissue[251]. Monzo and colleagues reported differential expression of 28 and 64 miRNAs in stage I and II colon cancer when compared with corresponding normal colon[252]. Recent scientific studies reported that miRNA-21 expression correlated with advanced colon cancer (Dukes' C and / or lymph node metastases). These results suggest that variable expression profiles of different miRNA molecules may be a potentially reliable method of diagnosing colon cancer and assessing its outcome. This is certainly a great advancement over

standard clinical staging of colorectal cancer, based on which further management is decided. Any human error on the part of the histo-pathologist in detecting micro metastases in the lymph nodes is liable to under stage the disease, resulting in inadequate treatment and poor outcome[253,254].

1.1.3.3. Proteomic level biomarkers:

Ultimately all cellular functions are performed and regulated by proteins. Thus, transcriptional profiling and genomics will never fully describe cell biology, even though in many cases such information can be used to predict cell behaviours. The discrepancy often noted however between these approaches and those which look directly at proteins and protein modifications result from the following:

1. Levels of mRNA do not always directly correlate with levels of proteins, as these can be influenced by many processes that may not be revealed by gene expression studies.
    a. Post transcriptional regulation, by various factors including, non coding RNAs; regulation of translation and post-translational modification of proteins
    b. Regulation of protein stability, degradation etc
2. Protein modification, in particular by altering phosphorylation can switch proteins, and therefore biological processes on or off, without any changes in mRNA or protein synthesis.
3. Protein dynamics, translocation and complex formation.

Proteomics is a field of biology directed at addressing these issues[255]. The study of proteomics is aimed towards identifying proteins within the organism. Application and ongoing development of new techniques are aimed to identify temporal changes in protein levels associated with normal tissue and tissue affected by tumorigenesis[256]. These techniques are described below:

- **Two-dimensional denaturing poly-acrylamide gel electrophoresis (2D-PAGE)**

The initial description of this technique dates back to 1975[257]. In its initial stages of development, separation and identification of target proteins in the first dimension, was carried out by iso-electric focusing (IEF) according to their respective iso-electric points. Thereafter, introduction of immobilized pH gradient (IPG) strips improved this technique in terms of reproducibility and capacity[258]. In the second dimension, proteins are then separated according to their molecular masses, mostly by (sodium dodecyl sulfatepolyacrylamide) gel electrophoresis. After the gel is stained with silver or Commassie blue the respective proteins are seen as spots or dots representing differences between different samples. Special software is then used to study the gels, in order to assess the differences in between proteomes. The potential of 2DE as a tool for identifying biomarkers has been greatly augmented with the development of mass spectrometry techniques (MS), where the later is used to identify the spots, representing individual protein components[166]. Despite the fact that 2D-PAGE is currently an important research tool in the study of proteomics, it has several limitations. Along with being a time consuming and labour intensive technique, membrane bound proteins are inadvertently under represented and poor solubility of a number of isolated protein molecules can pose a problem[259]. Again, quantification of protein components is based on respective staining signals, which can be inaccurate for proteins with very low concentrations. Inter gel variability prevents superimposition of images, which is partly solved by introducing difference-gel electrophoresis (DiGE), using highly sensitive fluorescent dyes[258,260]. Protein molecules are covalently bound to two different fluorescent dyes (Cy3 and Cy5), which are thereafter mixed and separated on 2D-PAGE. The gels from individual samples are scanned by different wavelengths to display the results, comparison of which, provides information on relative abundance of proteins in these samples[256]. Ambiguities involved in “spot” matching can thus be avoided, as one uses the same gel to separate spots. Many differentially expressed

proteins have subsequently been detected by DiGE and functional phenotypes identified, that might be an inducing factor for brain metastasis in breast cancer or nodal metastases in oesophageal cancer[261,262]. Although, DiGE is associated with a low throughput and limitations in analyzing insoluble proteins, perfect image super imposition for image analysis, which may be a problem while dealing with large number of samples in multistep 2DE, is less time consuming and relatively easier with DiGE[263,264].

- **Mass spectrometry**

Mass spectrometry (mass-to-charge ratio of gas-phase ions) is a valuable tool for identifying and quantification of proteins, defining protein interactions, post-translational modifications, N-terminal, C-terminal and de novo sequencing. In this technology, electrical charge to the molecules is conferred by an ion source. A mass analyzer thereafter separates these ions according to their m/z ratio, which are detected by a detector[265,266]. Ionization of proteins is achieved either by matrix-assisted laser desorption/ionization (MALDI) or electro spray ionization (ESI). MALDI uses a laser to vaporize and ionize the sample protein molecules in a matrix. In ESI, ionization is applied to a stream of liquid containing the sample, creating a stream of ions, which is then analysed by the mass spectrometer[173]. Some analyzers used most commonly are quadrupole (Q), ion-traps (for example, quadrupole ion trap or QIT; linear ion trap or LIT), time of flight (TOF) and Fourier transform mass analyzer (ion cyclotron resonance—FTICR). Several hybrid instruments combining capabilities of different analyzers were also recently developed (example, dual-cell quadrupole linear ion trap-orbitrap—dcQLT-orbitrap, LTQ-FTICR, etc.). MALDI is coupled with TOF mass analyzer, which accelerates ions through an electrical field of known strength[267,268]. Both of these techniques have been applied in identifying novel biomarkers. Metastatic cell lines were compared by using 2D-PAGE and MALDI-TOF. Vimentin, a promoter of tumour cell invasiveness, and diagnostic peptidome signatures from colorectal cancer were

identified using this tool[269,270]. Protein patterns comparing and predicting survival of patients with gliomas were detected Schwartz and colleagues[271]. Vydra *et al* reported three differentially expressed proteins in patients with breast cancer with distant metastases[272]. To reduce sample complexity and to deplete highly abundant more unimportant plasma proteins, several approaches have been used. The most widely used method is separating proteins by 2D-PAGE prior to MS. In addition, mass analyzers are often directly coupled with other methods like liquid chromatography (LC) and capillary electrophoresis (CE), which enables separation of protein molecules, such as multi-dimensional protein identification technology or Mud-Pit, based upon affinity columns enabling sequential cation exchange, and reverse-phase chromatography[173,188].

- **Surface-enhanced laser desorption/ionization (SELDI)**

This technique, a variation of MALDI (discussed above), uses a chemically coated plate (in place of a metal plate as in MALDI). The analyte solution is spotted onto this coated plate. Some proteins in the mixture bind to this chemical on the plate. Subsequent washing step removes the un-bound protein from the analyte. This works as a separation step. The matrix is then spotted on the plate, allowed to crystallize along with the bound peptide and irradiated with laser. Measurement of ionic mass is done in similar lines as discussed in MALDI. Selective coating of the SELDI plate can be done with antibodies to bind corresponding proteins[273-275]. Coupled to TOF spectrometers, this method has the capacity of producing proteomic fingerprints from biological samples. High throughput analyses are possible with this approach. However, its use in biomarkers discovery is limited due to low-resolution [276,277]. This approach has been used by Smith *et al* to assess response to neoadjuvant radio–chemotherapy in rectal cancer. They measured 14 differentially expressed proteins after 1 day and 2 days of treatment in their cohort and demonstrated that pathologic response to treatment could be predicted with 87.5% sensitivity and 80% specificity and also suggested that SELDI-TOF-MS are capable of detecting biomarkers and predict therapeutic response at an

individual level[278]. Other reports have used similar approaches in breast, ovarian, colon, liver and gastric cancers with promising results[164,170,279,280].

- **Tandem mass spectrometry (MS)**

Identification and sequencing of protein and / or peptide is possible by MS. It involves multiple steps of separation in time (ion traps) or in space. Most widely used techniques are collision-induced dissociation (CID), involving multiple collisions of protein ions with rare gas atoms, electron-capture dissociation (ECD) and electron-transfer dissociation (ETD), inducing fragmentation by transferring electrons to proteins, and infrared multi photon dissociation (IRMPD) by infrared laser[173,265]. MS based technology has enabled us to detect post translational modifications and dynamic changes of protein molecules. Relative abundance of target proteins in disease versus healthy tissue can be detected by labelling these respective proteins with stable isotopes. The differentially labelled samples are then separated on affinity columns or LC, followed by which they undergo analyses. Samples can be labelled chemically by several means; ICAT or Isotope coded affinity tag, where isotopic labelling of amino acid side chain is achieved, isobaric mass tags, such as tandem mass tags or TMT and isobaric tags for quantization (iTRAC and new label mTRAC) based on incorporating reporter groups with the same mass. Stable isotope labelling of amino acids in cell culture (SILAC) has been conducted *in vivo* and *in vitro* models[265,266,281,282]. Label free quantification methods such as spectral counting are applied to biological samples, where they are not combined, but measured separately[260].

- **Protein microarrays**

Protein microarrays enable global studies of protein expression, post translational modifications, and interaction in between two protein molecules and protein–ligand and protein–kinase interactions. Although high-throughput analyses is possible, there are some technical obstacles in the way of robust protein assay, such as selecting

appropriate methods for generating, purifying and immobilizing proteins, determining specificity for different antibodies and solving high sample variability[188,283,284]. Variable domains of single chain antibodies, in the form of “nano-bodies” were generated by some researchers, as a simple and cost effective approach for immobilizing proteins on microarray surface[285,286]. Reactive proteins are detected on chip by small molecule probes being labelled with fluorescent, magnetic, affinity, photochemical or radioisotope tags[287,288]. Label free detection method for protein interactions are being developed, such as surface plasmon resonance (SPR) and high-speed oblique-incidence reflectivity difference (OI-RD) microscope[289]. Laser-capture dissection or cell sorting techniques are being used to isolate single cell populations as protein microarrays are not cell-specific[290-293]. Assuring stability of immobilized proteins is a potential problem due to rapid decrease of their activity. The protein in situ array (PISA) solves this problem and seems to be promising for future development of protein microarrays. The method generates protein arrays directly from (PCR amplified) c-DNA, generated by cell-free synthesis of tagged proteins onto the tag-capturing surface[294,295]. Several types of protein arrays, such as analytical protein microarrays, antibody, reverse phase, functional and bead-based microarrays have been developed and extensively reviewed[283,284,288,296]. Advanced stage ovarian cancer has been shown to over-express phosphorylated extracellular-signal-regulated kinase unlike early stage disease, by the application of reverse phase microarrays[297]. Protein ultra microarray was used for the detection of IL-6 and PSA[298]. Important tumour marker antigens such as AFP, CEA, β-HCG, CA 125, CA 19-9 and CA 15-3 were detected by similar techniques[299]. In a study, conducted by Gulmann and colleagues, laser-capture micro dissection was applied to isolate tumour epithelium, stromal and normal cells from tissue samples obtained from patients with colorectal cancer. Reverse-phase protein microarrays was used to establish decreased levels of activated isoforms of proteins like p-ERK, p-p38 and p-JNK, found to be involved in MAPK-kinase signaling pathway, as discussed above[293].

- **Laser-capture micro-dissection (LCM):**

This computerized technique uses laser energy to isolate different cells of interest (normal, cancer, etc) from a tissue specimen, without denaturing them. However, given the limitations like extremely specialized level of technology and relatively low protein yield, its use for generalised cancer screening is not yet possible[300].

- **Bioinformatics in proteomic studies**

Comparing and determining protein molecular composition has become easier with the introduction of bio-informatic tools in proteomic research. Large data bases like Swiss–Prot and BlastPro have been created and large heterogeneous data sets can be compared against this database to gain information about new disease processes. New algorithms for image analysis of 2D-PAGE gels, MS analyses or high density protein microarrays have been developed within the last years[283,301,302].

- **Two dimensional difference gel electrophoresis (2-D DiGE):**

In this form of gel electrophoresis, up to three different protein samples are labelled with separate fluorescent dyes, the later differing in their respective excitation and emission wavelengths. After 2-D gel electrophoresis, protein abundance is noted by repeated scanning of the gel with excitation wavelength of one fluorescent dye after the other. In this technique, proteins from different samples are run on the same gel; hence the inter-gel variation as found with 2DE is substantially reduced.

A) Tissue Biomarkers:

The above mentioned techniques have been employed extensively by several researchers and numerous tissue biomarkers have been reported from time to time. Calgranulin B was found to be a prominent protein of uncertain functional significance in colon cancer tissue and in pre cancerous lesions as compared to their normal

counterparts by 2DE[303]. 2D-DiGE and MALDI was used to report differential expression of nearly 52 different proteins in mucosal specimens of colorectal cancer as compared to normal colonic mucosa[304]. Similar approach was used by another group to report differential expression of vimentin, cytokeratin, β–actin, heat shock protein 60, cathepsin D, RSP4, calreticulin, relaxin, APC, metastasis associated protein, annexins IV and V in colon cancer versus normal colon[305]. Locally advanced rectal tumour was investigated using narrow pH range 2-DE and MALDI and increased expression of following proteins were detected: tropomodulin, β-tubulin, calsenilin, annexin V, heat shock protein 42, DNA repair protein (RAD51L3), notch 2 protein homolog and keratin type I[306]. Nature and progression of colon cancer through different stages (normal mucosa → adenoma →Dukes' stages B to D →metastatic colon cancer in liver) of the disease was studied by 2D-PAGE and cDNA (complementary DNA) microarray analysis by Kwong and colleagues. Parallel assessment of the genes and resultant proteins was done in different stages by the authors[307]. Colon and ovarian cancer cell lines were compared in a report using multistep genomic and proteomic analyses. Villin and moesin were identified as markers for colon and ovarian cancer cells respectively. The group used techniques like cDNA microarray, affymetrix oligonucleotide chip analysis (corroborating transcript levels), reverse phase protein micro-array (quantification of proteins) and tissue micro-array[308]. Cancer therapeutics has also been influenced by these techniques used for investigating cancer tissue. A pelvic mass of uncertain origin was analysed by 2-DE, compared against 2-DE analyses database of high grade colon and ovarian malignancies, and was found to be of colonic origin and treated accordingly[309]. Possibilities of recurrence in Dukes' B colon cancer were predicted in a report with 78% accuracy by using affymetrix U133a Gene Chip with nearly 22 thousand transcripts. If similar studies could be validated in larger groups, published data with higher levels of accuracy can be helpful in preventing cancer relapse through implementing adjuvant therapy[310].

B) Serum Biomarkers:

An ideal biomarker is detected in a readily available body fluid such as serum, differentiates healthy individuals from cancer cases and predicts the state and course of the disease[311]. Although extensive research is ongoing in the field of serum biomarkers for screening / detecting colon cancer and some progress has been made in serum proteomics, results so far are largely disappointing[312].

Carcino-embryonic antigen (CEA) is routinely used as a serum biomarker and an independent prognostic measure for colorectal cancer in today's clinical practice. It is a highly glycosylated protein and has been described for the first time in 1965[313,314]. Higher levels of serum CEA has been shown to correlate well with poorly differentiated and advanced stage of tumour[315]. Preoperative CEA levels have also been shown to predict post operative survival in patients with colorectal cancer[316]. However, based on its diagnostic sensitivity (30-40%) and specificity (87%), Fletcher and colleagues calculated that there would be 250 false positive cases for every cancer patient detected by CEA estimation and 60% cancers will be missed[317]. Many metastatic cases of colon cancer in livers (67%) and lungs (94%) have been missed when CEA was used as the sole diagnostic tool[318,319]. Again, high CEA levels in response to non-malignant causes like hepatitis, chronic obstructive pulmonary disease, pancreatitis and inflammatory bowel disease, persistently high levels in smokers (19%) and high serum levels even without any evidence of cancer recurrence puts its reliability and sensitivity under question[320-322]. These values draw a limitation to its use as a diagnostic marker; however its stand in colorectal cancer surveillance, after curative surgery is still widely accepted[315]. Serum levels are expected to decrease following curative resection and increase in recurrent cancer. Given its low sensitivity and specificity in Dukes' A and B colorectal cancer, the NIH (National Institute of Health) recommends that it should not be used to screen for early colon cancer[323]. Macrophage migration inhibitory peptide (MIF) is elevated in serum from colon cancer patients. Its sensitivity in early colon cancer detection is higher than CEA (47.3% vs 29.5%)[324]. The sialylated Lewis-a

antigen, CA 19-9 (Carbohydrate antigen 19-9) is currently used as a serum biomarker for pancreatic carcinoma. Its role as serum biomarker in colon cancer has also been tried, but the sensitivity is even lower as compared to that of CEA. Other similar antigens like CA50, CA242, CA M26, CA72-4, CA195, CA M26 and CA M43 have also been investigated from similar perspective in colon cancer, but sensitivities were found to be very low[325-327]. Serum levels of Matrix metallo proteinases (MMPs 7 and 9)[328,329], prolactin[330], G-CSF (granulocyte colony stimulating factor)[331] and M-CSF (macrophage colony stimulating factor) has been found to be elevated in colon cancer with varying sensitivity and specificity. The later has also been reported to indicate lymph node metastases in colon cancer[332]. Serum levels of MMP7 and laminin have been reported to be independent prognostic markers for colon cancer[328,333].

TIMP-1 (Tissue inhibitor of metalloproteinase type 1) is a glycoprotein, which is found to be over expressed in serum from patients with colorectal cancer as compared to healthy individuals and those with colonic adenoma, inflammatory bowel disease or primary breast cancer[334-336]. Its role as an independent serological prognostic marker in preoperative analysis has been reported[337-339]. However, its expression is not restricted to colon cancer alone, but also present in other conditions like cancers of stomach[340], urinary bladder[341], lung[342] and ovary[343].

An equal or better diagnostic accuracy for predicting colon cancer, as compared to that of CEA, by serum levels of NNMT (nicotinamide N-methyl transferase) and PSME3 (proteasome activator complex subunit) has been reported by two studies [344,345]. Wu and colleagues reported that serum levels of CRMP-2 (collapsing response mediator protein) have a better sensitivity but poorer specificity as compared to that of CEA in diagnosing colorectal cancer[346].

Albrethsen and colleagues used SELDI-TOF to demonstrate up-regulation of HNP (human neutrophil peptide) – 1, 2 and 3 (alternatively called α-defensin-1, 2 and 3 respectively), both in serum of colon cancer patients and mucosa from cancer specimens as compared to healthy individuals[347,348]. One subsequent study reported

that apo-lipoprotein C1, C3a-desArg, transferrin and α1-antitrypsin as markers of high diagnostic potential in sera from colon cancer patients by using SELDI and ELISA (enzyme linked immune sorbent assay)[349]. Habermann *et al* used ELISA (detecting C3a-desArg levels in serum) to diagnose colon cancer with nearly 96 % specificity and 97 % sensitivity. However, according to this later report, C3a-desArg levels were elevated in 86% serum samples only[349,350]. Using SELDI and highly sophisticated bioinformatics tool on serum samples from colorectal cancer patients of different stages (Dukes' A to D), patients with colon adenomas and healthy individuals, Yu *et al* reported that patients with adenomas and cancer can be differentiated from the rest of the cohort with 89% sensitivity, 83% specificity and 89% predictive value[351].

C) Stool Biomarkers:

Based on the recommendations by the American Gastroenterological Association, faecal occult blood testing has been introduced as a widely accepted screening test for colon cancer. It has been shown to reduce mortality in this disease by early detection[352]. However, false negative cases may range up to 20-30% as reported in the literature due to its poor sensitivity and due to the fact that all cancer may not bleed[353]. Again, false positive cases may be substantially high if the diagnosis is based solely on FOBT, given the numerous non-malignant causes of gastrointestinal bleed. Annual re-testing for FOBT currently detects only 25-50% colon cancers and 10% adenomas[354]. Although separating and amplifying abnormal mutated (cancer causing) DNAs from normal human and bacterial DNAs possess a potential technical challenge and has been compared with "finding a needle in the haystack", higher integrity of DNA fragments in the stool of colon cancer patients (as compared to normal patients), possibly due to inhibition of apoptosis and auto digestion in cancer, is potentially advantageous[355,356]. Colon cancers continuously exfoliate cells in the stool. Diagnosing cancer by detecting mutated forms of DNAs from these colonocytes in a potential future non-invasive mode of cancer screening and is still under investigation. The screening targets most

commonly tried with stool are mutated *k-ras*, *p53* and *BAT-26*[213,357]. Faecal DNA testing was also found to yield better results, with diagnosing colon cancer, when compared against FOBT in a large cohort of more than 4000 average risk adults of age more than 50 years. In this study by Imperiale *et al*, sensitivity of fecal DNA testing was far better than FOBT (52% *vs* 13% respectively) for diagnosing invasive colon cancer and just slightly better (15% *vs* 11%) in detecting adenomas with low or moderate grade dysplasia. The specificity of faecal DNA testing and FOBT were almost found to be same (94% *vs* 95% respectively)[358]. However, several questions put forward by Steven Woolf in his editorial makes us think about the applicability of DNA analysis from faeces as the sole screening test for colon cancer[354]. Faecal calprotectin has been reported to be elevated in colon cancer and inflammatory bowel disease[359,360]. But a lower specificity as compared to FOBT (63% *vs* 97%) makes it a poor biomarker for colon cancer[361].

1.1.3.4. Multiple markers in colon cancer:

From our previous discussion, it is apparent that the use of single proteins as biomarkers for a certain cancer is often non-specific. Therefore the use of multiple markers has been tried by several researchers. Serum levels of five protein panel: TRAIL receptor-2 (TRAIL R-2), spondin-2, Reg IV, macrophage inhibitory cytokine 1 (MIC1) and tumour necrosis factor receptor superfamily member 6B (DcR3) has been used as biomarker for colon cancer in conjunction and has been shown to have better sensitivity and specificity than CEA used alone[362]. Similarly, serum levels of two protein panel, CEA and CRMP-2, when used together, was found to have better sensitivity (77%) and specificity (95%) than CEA (sensitivity and specificity of 43% and 87% respectively)[346]. M2-pyruvate kinase alone has a sensitivity of 48-58% and specificity of 90-95% as a serum biomarker in colon cancer. However, its sensitivity increases when used in conjunction with CEA, without a fall in its specificity[363,364]. Use of gene microarray in diagnosing colon cancer, by detecting a panel of 5 differentially regulated

genes, *BANK1(B-cell scaffold protein with ankyrin repeats 1), BCNP1 (B-cell novel protein 1), CDS(Cytidine deaminase), MGC20553(FERM domain containing 3)* and *MS4A1 (membrane spanning 4-domains, subfamily A, member1)*, from the total extracted pool of genes from peripheral white blood cells, was found to be very sensitive (88-94%) but with lower specificity (64-77%) as compared to that of CEA[365]. Similar combination strategies applied to stool biomarkers was found to be advantageous. Although, presence of mutant *k-ras* in less than 50% cases of colon cancer is responsible for its low sensitivity as a single stool based biomarker[366], combination of the other two targets (*p53* and *BAT-26*) detected 71% colon cancer patients in a cohort[357]. When mutated *APC* and long DNA (a marker for non-apoptosis) were combined with the above, nearly 91% cases with colon cancer and 82% cases with adenomas (less than 1 cm) were detected. On removing *k-ras* from this 5-tag panel, the sensitivity dropped slightly but the specificity rose to 100%[367]. As discussed before, faecal DNA testing has better sensitivity and comparable specificity to FOBT in detecting colon cancer[358]. Combination of FOBT and faecal DNA in large cohorts may be important in increasing the sensitivity and specificity of screening for colon cancer.

In this context, Professor Zena Werb has rightly said, *"As attractive as the concept of a single marker is, it is now crucial to consider what such markers really mark. These molecules, which often lack tumor-specific function, may help tumor cells maintain their characteristics in a specific niche, or they may be expressed because of another gene turned on in the tumor or because of altered adhesion. So it is not hard to imagine that the markers may not be essential or constant, or that they may be expressed in normal tissues. For understanding tumorigenesis and creating therapies, a spectrum of markers may be the answer"*[368].

In a study with normal appearing colonic mucosa from patients with colonic adenomatous polyps (and risk of colon cancer), higher expression of CD166, EpCAM and CD44 was found to be associated with increased EGFR (Epidermal growth factor

receptor, thought to promote proliferation in cancer cells), number of polyps found and age of the patients [369].

As already mentioned, origin of colon cancer can be explained by 2 popular models, stochastic model and cancer stem cell model. The first strong support for the stem cell theory in colon cancer comes, from two recent studies by O'Brien *et al* [38] and Ricci-Vitiani *et al* [37] . Based on their xeno-transplantation experiments, these two groups reported that CD133 is a marker for colon CSCs [37,38]. However, a more comprehensive study performed in similar lines by Dalerba *et al* revealed that only the $EpCAM^{Rich}$ / $CD44^{+}$ / $CD166^{+}$ fraction of cells represents the proliferative segment of a colon cancer mass, in contrast to the cells expressing only CD133, which was found to be a variable indicator for CSCs by these authors. They also reported that CD49f was a more consistent marker for most cells expressing CD44. However, the authors failed to establish the specificity of this combinatorial protein expression to be a unique signature of a colon cancer stem cell [39].

1.1.4. Toponome imaging system:

The different levels at which the functionality of a cell is determined are genomic, transcriptional, and at proteomic levels. Protein networks within and on the surface of cells are highly organized, both timely and spatially, determining specific cellular functions at different time points. The living cell has been described as a *"protein co-compartmentalization machine",* where *"each protein must be at the right time, at the right location and at the right concentration in a cell to interact with other proteins"*[194] . If we are to believe that every cellular function is represented by a specific *in situ* pattern of protein network in a cell, detecting such pattern can be of pivotal importance in understanding the cell's physiological and / or pathological behaviour[193].

The ultimate goal of cell biologists is to decipher the chain of life from molecules through cells to tissues and finally to organisms. In order to clearly understand how cells work, robust knowledge regarding identities, structures, functions of individual molecules and

their interaction with each other is therefore imperative[370]. Although information on molecular structure and function of individual proteins is steadily on the increase, our knowledge regarding formation of protein clusters at sub cellular level, interaction of these clusters as networks to generate functions and finally reassembling of protein molecules to form new clusters over time, generating new functional networks, is still at its infancy[371].

Formation of protein networks is hugely determined by the cell and its interactions with surrounding cells and inter cellular matrix. Homogenisation procedures notoriously destroy the 3D information of cellular distribution of proteins in a toponome and also the distribution of cells in its natural habitat. If we are to believe that in order to form a functional protein cluster, each component protein must be present at the right time, right place and in the right amount, a system which breaks down tissues before analysing the components (homogenisation techniques) can only provide quantitative information, but fails to confirm whether these proteins were at the right place or at the right time to react with the right protein. Again, from our discussion above regarding normal colonic stem cells, we know that when stem cells are removed from their protective niche in the crypts, it affects their 'stem-ness'. This may indicate that a majority of cellular functions are also related to cell – cell and / or cell – matrix interaction. Thus, a system which reads a part or whole of the protein network, yet conserves the tissue architecture is essential to understand not only the protein network pattern of the 'cell of interest', but also its relationship to the matrix and surrounding cluster of cells in the tissue. Single cell 3D proteomics technology thus analyses functional protein molecular networks, which are undetected by large scale expression profiles based on tissue homogenates. Again, from our understanding of cell biology, we know that even with similar expression levels, different molecular distribution patterns are responsible for entirely different characters in the same cell.

For example, let us assume that 'Protein A' is responsible for cell adhesion. After being formed in the cell, it is carried by a 'carrier Protein B' to the cell surface, which is an

essential stage for Protein A to express its function. The inability of transporting protein A from cytoplasm to the cell surface, due to non-formation of Protein B may thus be responsible for increased deposition of this molecular component in the cytoplasm and may explain a different character like decreased adhesion or increased mobility of this cell resulting into its invasive behaviour (example in a malignant condition) as opposed to the normal state where this protein is evenly distributed in the same quantity on cell membrane. In this example, homogenization techniques would identify no difference between diseased and normal cells as the relative molecular quantities are constant in either cell. Only a technique which reads the protein network code in intact tissue sections would explain this difference in cellular behaviour by localizing Protein A.

In the past, the human genome project had revealed the code for all proteins, among other things. The next big challenge was to understand how these proteins interact in a cell with other proteins in a timely and spatial manner. Development of new technologies with the capacity to co-localize a huge number of proteins in one biologically intact section in order to reveal these functional protein patterns (assemblies or clusters of different proteins, in the form of a protein distribution map, also considered as the protein code or the "biological code") generating concrete cell functions. It is now being possible to explore biological structures through developing techniques, allowing us to study supra molecular aspects of protein network from the level of the tissue down to the sub-cellular compartments of single cells.

A technology has recently been described by Schubert and colleagues, which enables us to systematically address such issues in morphologically intact tissue sections by using an automated immune-fluorescence imaging technology called TIS (Toponome Imaging System) [372,373]. TIS has the capacity to identify functional and hierarchical inter relationships of proteins on a large scale in a single cell in intact tissue sections. This level of protein organization or the entirety of protein network has been referred to as the 'toponome' (from the ancient greek words 'topos' = place, position; and 'nomos' = law)[193].

In order to address the internal structure of such complex sub cellular *in situ* protein networks, the huge number of protein molecules that needs to be co localized in a single tissue section can significantly exceed the number of available wavelengths in traditional fluorescence microscopy for their respective spectral isolation. This problem was solved by demonstrating a basic fluorescence method, breaking this spectral limit, by application of a cyclical one/two-dye imaging system[374].

TIS was applied to solve the "Riddle of Myogenic Stem Cells" with a panel of 19 different protein marker tags[375]. By using protein co-localization techniques, the authors reported that endothelial cells from surrounding micro-vessels migrated to the damaged portion of the muscle fibres and formed new myotubes. Furthermore, the authors identified unique protein combinations (using their randomly selected panel of 19 marker tags), for myogenic stem cells, newly formed myotubes, mature endothelial cells and muscle fibres. These findings pointed towards the possibility of trans-differentiation of mature endothelial cells into newly formed myotubes by selective protein switch on / off mechanism on cell surface (CD36 switched off in endothelial cells to form myogenic stem cells, with expression of CD56, the later being an isoform of NCAM and putative marker for mononuclear myogenic stem cells). These findings were supported later by alternative approaches by De Angelis *et al* (Cossu Group) [376].

In a subsequent report, TIS was used to investigate human muscle tissue, affected by Polymyositis, an inflammatory condition. The researchers applied a panel of 20 different cell surface receptor proteins, which were co-mapped simultaneously and were obviously mismatched in sections of muscle tissue (from chronic Polymyositis) with T lymphocytes invading the muscle cells. This experiment suggested that abnormal combinatorial surface protein organization may be selective for pathological conditions, for example inflammation in Polymyositis. It also demonstrated that the T cells through their migration from peri-vascular space to the muscles fibres (different stages in this migration were; *a]* peri vascular space, *b]* inside endomysium in contact to basal lamina cylinder and *c]* pacemaker T cells inside basal lamina cylinder) had different protein

combinations on their cell surface, which were interestingly specific for the micro environment the T cells were subjected to [377]. Based upon these initial findings, this same principle was further used to develop a robot assisted automated immune fluorescence technology which has the capacity of co-localizing a quasi random number of different molecules in one tissue section, called TIS[372,373].

In this method, dye-conjugated marker molecules (tags) are applied on a fixed biological sample. As soon as the first molecular component is labelled with the respective tag (post incubation stage), the corresponding fluorescence signal is registered as an image. Following this the fluorescence is bleached (post bleaching stage) and the first incubation–imaging–bleaching cycle is completed. Signals (fluorescence images) are also recorded following the bleaching step, when the tissue is expected to have minimal or no residual fluorescence signal. Phase contrast images are also recorded at the respective post incubation and post bleaching stages. This cycle is repeated over and over again with individual markers specific for other molecular components in respective cycles, each of which ends with the corresponding bleaching step. It has been demonstrated by the authors, using TIS experiments on sections from human skin, that this system has the capacity to successfully complete at least 100 such repeated incubation-imaging-bleaching cycles (RIIBC) on one biological sample leading to high-dimensional structure associated, combinatorial molecular data sets, on every data point of this sample, yielding a power of combinatorial molecular discrimination (PCMD) of at least 2^{100} per data point[372].

In practice, with traditional multi-channel fluorescence microscopy, at best signals from four or lower number of different tags can be identified separately. This is grossly insufficient to identify and explore large molecular networks in an entire biological structure[371]. In contrast, by using TIS, multiple molecular components can be simultaneously co-localized at a sub-cellular level. This enables us to study the functional architecture of a tissue on a larger scale and also provides us with unique

information regarding which proteins are inter-related ("mapping togetherness") and which proteins are not in a given physiological or pathological condition of the tissue[378].

In the protein co-localization experiments with TIS, images obtained from RIIBC, are aligned and are processed by subtracting background (false) signals. Protein signal combinations per pixel are identified in this data set. Thus every fluorescent image for every visual field is a vectorial representation for the respective protein molecular expression. Signal combinations or combination of these vectors per pixel are expressed either as relative grey value profile or following introduction of a threshold for each signal, as a combinatorial binary code. All these signal combinations (depicting protein clusters) are assembled as a data set known as "Toponome map". Such a data set can be used to visualize the location of a given protein cluster (combinatorial molecular phenotype or CMP) in groups of pixels or voxels, on the background of the section of tissue studied[195,373,379]. This imaging can be in 2D or in 3D[380]. Furthermore, these data sets can be archived for future reference and analyses. Needless to mention, thorough calibration experiments with all these tags needs to precede any such protein co-localization experiment to assure the specificity and selectivity of each individual tag used[371].

Co-mapping of a large number of different CMPs in one sample can be easily done by introducing a threshold grey value for each fluorescent signal. This threshold grey value is determined by the background fluorescence or tissue auto-fluorescence. In order to achieve this, at the beginning of a TIS protein co-localization experiment, few subsequent tags (number of such tags is determined by level of tissue auto-fluorescence and optimisation steps) are replaced by PBS (phosphate buffer solution) and RIIB cycles are allowed to run. In these cycles, the tissue is bleached without any new fluorescent tag added to the tissue. The fluorescent signal (post incubation image) obtained with the last PBS run is considered equivalent to the background fluorescence and can be used as the threshold grey value for the same experiment.

Location of each protein is detected by presence of signals within a CMP, mapped as tag absent (=0) or present (=1), depending on whether the value for the fluorescence signal is below or above the threshold signal, respectively (=1 bit information per protein at a pixel / voxel). CMP motifs thus defined as a group of CMPs possessing unique common features. These motifs are defined by the assembly's lead proteins (L = 1), absent proteins (A = anti-colocated, 0) and wild card proteins (W or variably expressed i.e expressed in some CMPs and absent in others). This three symbol code (L, A and W) can be used effectively to describe motifs of protein clusters unique to certain interesting regions in a tissue section (example glandular cells versus stromal cells) or a section in a certain disease state, (for example, present in cancer but not in normal condition for the same tissue) [195,373,379]. The use of photonic microscopy and use of increased number of tags further refines this technology. TIS has been used by several other researchers since its first publication[374] in 1990 by Schubert and colleagues[379-387].

In vitro TIS experiments using Rhabdomyosarcoma cells (from cell line TE671) with a panel of 23 cell surface proteins, revealed the significance of identifying and inhibition of lead proteins in functional clusters[372]. The authors described different stages of disease, where a resting (spherical) cell transforms in invasive (exploratory) cell after developing cell extensions. Interestingly, protein co-localization by TIS revealed identical arrangement of 5 CMPs formed of only 6 proteins (from the above panel of 23 marker tags) located on all these cell extensions with the lead protein being CD13 (Aminopeptidase N) in all of these combinations. By selectively inhibiting CD13, the authors demonstrated that the exploratory cells lost their cell extensions and transformed back to resting spherical cells. Spherical cells on the other hand, failed to develop cell extensions; thus could not reach the invasive phase. This finding was exciting, as it pointed towards the possibility of downgrading Rhabdomyosarcoma by selective inhibition of lead proteins in functionally important protein clusters and its potential applicability in other cancers. This experiment also revealed a new cell polarization function (APOCON model; Aminopeptidase Polarization Control Network

Model) for CD13, which was unknown through other scientific techniques. Similar experiments with Amyotrophic Lateral Sclerosis cells lines revealed FCγRIII (FCγ Receptor III) protein as the lead protein in the defective nerve fibres[372]. Later on, using FCγ knock-out mouse models, Mohamed and colleagues confirmed the significance of blocking /inhibition of this target protein to control disease[388].

In an ongoing research project, Berndt and colleagues[389] investigated the role of T-cell mediated immune response in colon cancer. They reported that T-cells (mostly CD8+ as compared to CD4+) were grossly increased in lamina propria in colon cancer with increased expression of NF−κβ and CD29 (increase proliferation and decrease apoptosis) in these CD8+ cells. As an inference, the authors raised the possibility of up-regulation of host mucosal auto-defence mechanism due to cancer, which can potentially be exploited for cancer immunotherapy. Experiments with a single section of prostate cancer by Schubert and colleagues demonstrated the capacity of TIS in identifying motifs for cells demonstrating features of PIN (Progressive Intra epithelial Neoplasia)[379].

To summarize, TIS is a significant advancement towards this quest for reading the "biological code" of a section of tissue from human body, cell aggregates from cultures or body fluids. This system has the capacity to co-localize large number of molecular cell components independently in the same cell, within a morphologically intact tissue section, in the same experiment, under similar conditions, presenting the data as a coloured mosaic, which can stored as a database and compared with different patterns from other experiments in a systematic way, pixel by pixel[195]. This technology uses large molecular libraries of antibodies, peptides, lectins or different ligands to tag and localise corresponding molecular components in a specially fixed tissue section, by means of photonic signals (fluorescence). This fully automated novel technique is based upon the principle of diffusion kinetics within defined sub-cellular volumes. It detects and registers each individual molecule in a cell as a spatial signal map and aligns them relative to

other molecular signals from the same cellular compartment. The net result is expressed as a toponomic map of the cell(s) of interest. Comparing toponomes from same cell types in different pathological states of a tissue can give us disease specific molecular configuration of the same protein molecules present in normal physiological conditions, explaining their differential behaviour like invasion, metastases, etc [193].

CHAPTER 2

1.2. METHODS AND MATERIALS.

1.2.1. Specimen source:

Sections of colon cancer and corresponding normal colon tissue were collected from the same patient, diagnosed with Dukes' C1 non-mucinous adeno carcinoma (T4, N2, M0) involving the descending colon, with macroscopic secondary spread to lymph nodes and omentum, without any prior radio or chemotherapy, but had adjuvant treatment following the surgery. The cancer and normal tissues were extracted from the colectomy specimen, from the surface of tumour mass and normal looking colonic mucosa 10 cm distant to the visible tumour margin respectively. Experimental procedures with these tissues were performed on the basis of written consent from this patient.

1.2.2. Tissue preparation:

Tissues were immediately fixed in 4% para-formaldehyde (PBS) solution after collection. Following overnight cryo protection in 20% sucrose solution at 4^0C, tissues were embedded in OCT blocks and stored at -80^0C (minus 80 degrees) until preparation of frozen sections.

1.2.3. Preparation of cover slips for experiments with TIS:

Following steps were followed to prepare coverslips in-house for the tissue sections for our TIS experiments, in line with protocols described before[373]:

1) 5 millilitres of 3-aminopropyl-triethoxysilane (Fluka, cat no. 09324, kindly supplied by Dr. Walter Shubert's lab, Magdeburg, Germany) was mixed with 250 millilitres of acetone to prepare "silane solution".

2) Coverslips (24 x 60 mm, thickness; 0.13 - 0.16 mm, vwr cat no. 16004-096) were placed on a coverslip holder, placed in a glass chamber filled with freshly prepared “silane solution” and incubated for 60 minutes at room temperature.

3) The coverslips is thereafter washed by placing the coverslip holder in two separate glass chambers one after the other (each for 10 minutes) with clear distilled water.

4) Coverslips are now baked at 60^0 C for 12 hours. Dried coverslips are stored at 4^0C for a maximum period of 4 months.

5) Defective coverslips: Some of the coverslips prepared in the above process developed some dried water drops. These coverslips were discarded.

1.2.4. Preparation of tissue sections:

5 micrometre sections were cut from each block in a cryotome and every alternate section were placed on special coverslips (numbered in sequence) preparation of which has been discussed above. The temperature of the cryotome was set at minus 18 degree centigrade. These coverslips are attached to slides by OCT (tissue tech) and stored in slide holders at minus 80^0 C freezer for future use. The other alternate set of sections is collected on normal glass slides for Haematoxylin and Eosin stain. Thus for every tissue section for TIS, we had two adjacent sections (one immediately before and one immediately after) for H&E, numbered and stored in minus 80^0 C.

1.2.5. Antibody library:

We used a library of 20 different antibodies to identify the corresponding proteins in the cancer, normal and omentum tissue sections. All the antibodies (except DAPI or 4,6’-diamidino-2-phenylindole) were directly conjugated with FITC (fluoresceine-iso-thio-cyanate). Of these, seven recognised nuclear markers, twelve recognised cell membrane proteins and the remainder membrane and cytoplasmic markers. The list of the tags used and their respective technical details are summarized below:

Antibody (tag)	Concentration	Dilution	Company	Catalogue no.
DAPI	1.5µg/ml	1:150	Vector Lab. Burlingame, USA	H-1200
PCNA Rabbit Polyclonal IgG	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-7907 FITC
Bcl-2 Mouse, Monoclonal IgG_1	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-509 FITC
Bax Mouse Monoclonal, IgG_{2B}	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-7480 FITC
Cyclin A Rabbit Polyclonal IgG	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-751 FITC
Cyclin D1 Mouse Monoclonal, IgG_{2B}	0.2mg/ml	6:100	Santa Cruz Biotechnology, USA	sc-8396 FITC
Ki67 Monoclonal Antibody	0.1mg/ml	6:100	AbD Serotec , UK	HCA006F
CEA Mouse Monoclonal IgG_{2B}	0.2mg/ml	1:50	Santa Cruz Biotechnology, USA	sc-59875 FITC
CD44/HCAM Mouse, Monoclonal IgG_1	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-7297 FITC
CK-19 Mouse, Monoclonal IgG_1	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-53003 FITC
CK-20 Goat Polyclonal, IgG	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-17113 FITC special custom
CD29 Mouse, Monoclonal IgG_1	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-9970 FITC
CD36 Mouse Monoclonal, IgM	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-7309 FITC
CD57 Mouse Monoclonal, IgM	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-6261 FITC
CD133 Mouse Monoclonal, IgG_1	0.1mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-130127 FITC, special custom
Laminin α-1 Rabbit Polyclonal IgG	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-5582 FITC, special custom

CD166 Mouse, Monoclonal IgG_1	0.1mg/ml	1:5	AbD Serotec , UK	MCA1926F
EpCAM Mouse, Monoclonal IgG_1	0.1mg/ml	1:25	AbD Serotec , UK	MCA1870F
NCAM/CD56 Mouse, Monoclonal IgG_1	0.2mg/ml	1:5	Santa Cruz Biotechnology, USA	sc-7326 FITC
Muc-1 Mouse Monoclonal, IgG_3	0.1mg/ml	1:5	AbD Serotec , UK	MCA1742F
Muc-2 Mouse, Monoclonal IgG_1	0.1mg/ml	1:50	AbD Serotec , UK	MCA1743F, special custom

1.2.6. Rationale behind using different antibodies:

We selected a random set of markers (Table 1 at the end of manuscript), related to colon cancer for this initial optimising experiment. The rationale behind selecting cancer stem cell related antibody targets like CD133, CD166, EpCAM and CD44 has already been discussed in "introduction" above. A potential limitation in using more extensive range of markers was unavailability of most antibodies in the conjugated form and relative paucity of funds for this initial project. Here we would discuss the rationale behind using the rest of the antibodies in our experiment.

- **Nuclear markers:** Proliferating cell nuclear antigen (PCNA), Cyclin A, Cyclin D1 and Ki67 are markers for cell proliferation and are thought to be associated with metastasis and poor prognosis in colon cancer[390,391]. As previously discussed, aneuploidy is commonly found in colon cancer. Cyclin A expression in nuclei has been found to correlate well with colon cancers with aneuploidy[245]. The nuclear protein, Bax (Bcl2-associated X protein), is a known inducer of cell death. Loss of this protein imparts the cells with resistance to apoptosis[392].
- **CD29 (Integrin β1):** It is a cell adhesion molecule, expressed in high concentrations on the epithelial cell surfaces and is downregulated in advanced colorectal cancer along with other adhesion molecules like E-Cadherin. DNAzymes

targeting *Integrin $\alpha 1$* have been shown to down regulate neovascularisation and several other key tumourigenic processes in colon cancer both *in vivo* and *in vitro*[393,394].

- **CEA (Carcino-embryonic antigen):** It is expressed in cytoplasm and cell membrane of normal columnar epithelial cells and goblet cells in colon, but levels are increased in many patients with colorectal cancer particularly those with hepatic metastases; CEA is used to follow up colon cancer patients in clinical practice[395-397]. Its importance has been discussed in the section of Biomarkers in "introduction".
- **Muc-1 and Muc-2.** Mucins are epithelial glycoproteins and can either be of transmembrane type (eg. Mucin-1 or Muc-1) or secreted gel forming type (eg. Mucin-2 or Muc-2). Muc-2 is over-expressed only in mucinous adenocarcinoma, but not in non-mucinous types. Though still under debate, the current view regarding colon cancer is that, up-regulation of Muc-1 and down regulation of Muc-2 correlate with advanced disease and distant metastases[398,399].
- **Anatomical markers:** We used DAPI[400] to define the nuclei. Relevant tissue anatomy was marked by Laminin[401], CD36[402], NCAM (neural cell adhesion molecule)[403], CD57 (HNK-1)[404-406] and cytokeratins (CK-19 and CK-20). Along with anatomical information about intracellular cytoskeleton, CK-20, a known marker of enterocyte maturity[407], is also known to be over-expressed in colon adenocarcinoma as compared to normal colon[408,409].

1.2.7. Toponome imaging system (TIS):

Detailed description of TIS^{TM}, purchased from ToposNomos Ltd. (Munich, Germany; www.toposnomos.com), has been published previously[195,372,373,379]. Briefly, TIS is a highly standardized robot imaging system for use, among other things, in the human toponome project[48]. TIS comprises of an inverse epi-fluorescence microscope, a cooled CCD camera, and a pipetting robot running many repetitive cycles (up to > 100) of tag-dye incubation, fluorescence imaging and bleaching on the microscope stage to

generate a large number of protein location images of the same cell/tissue structure (Figure 7).

The cover slip with respective tissue section was mounted on the stage of the inverted wide-field (Zeiss) microscope implemented as part of the TIS^{TM} imaging robot. We used fluorescence filter for FITC at position one (emission 520/35 BP, excitation 472/30 BP, beam splitter 495LP), originally provided from AHF analysentechnik (Tübingen, Germany, catalogue number F36-525) and installed DAPI filter (Filter set 49, catalogue number 488049-9901-000, emission 445/50 BP, excitation G 365, beam splitter 395 FT) to visualize the nuclei. For each tissue section, the experiment consisted of 25 repetitive-incubation-imaging-bleaching (RIIB) cycles in line with the protocol previously described[372,373]. The first four cycles consisted of PBS (described as PBS 1, 2, 3 and 4). For ruling out false staining, we used a fifth cycle with PBS at an intermediate position in the run. The temperature of the robotic system was maintained at 20°C. In each cycle, phase contrast (at the beginning and end of the cycle, named IP and OP images respectively) and fluorescence images (both before and after tag addition, called IF and OF images respectively) were taken by a highly sensitive, cooled (- 20°C) CCD camera, fitted to the side panel of the microscope. The magnification was adjusted to 63 times and light intensity for the microscope bulb was set to 3.2 V. Shutter opening and closing times were co-ordinated with the bleaching, end of incubation and washing times by the TIS^{TM} software running the system. For all the antibodies, except DAPI, shutter opening time was programmed to 600 ms for the phase image and 1000 ms for the fluorescent image. For DAPI, these times were 2000 ms and 100 ms respectively. Almost all the antibody tags used in our experiments have been used in diagnostic medicine and/or cancer research in the past. Positive signals are therefore expected to mark the positions of the target proteins.

Figure 7: This figure demonstrates the toponome imaging system used for performing these experiments with colon cancer and normal colon. Numbered arrows are used to identify the important components:

1. Buffer tray covered by thin aluminium foil, to be pierced by the pipette tip.
2. Antibody tray, to be filled with antibody solutions of specified dilutions.
3. Pipette tip attached to the end of the robot arm.
4. Camera attached to the side port.
5. Microscope
6. Glass slide for the tissue section.
7. Touch panel control for microscope.

1.2.8. Optical parameters and magnification:

We used 63 times objective without any additional magnification from the microscope. The CCD sensor has a nominal pixel number of 1024 x 1024. The pixel size is 13.0 x 13.0 micrometer. So the sensitive area is 13.3 x 13.3 millimetres. One pixel in the image equates to 206 x 206 nanometres. When viewed through the eye-piece (10x magnification) the net magnification with x63 objective will be 630 times. However, the

camera attached to the side-port does not use the eye-piece. Hence magnification of the images visualized on the computer screen is as follows:

$$\frac{\text{Diagonal of monitor (225mm)}}{\text{Diagonal of CCD sensor (19.15mm)}} \times \text{Objective (x63)} = 740.21 \text{ times}$$

Thus the images shown or are analyzed are approximately 740 times magnified. All analyses are done on these images; hence the slight difference in magnification, when visualized through the eye-piece has no implications on the final results.

1.2.9. Haematoxylin and Eosin (H&E) staining of slides:

Following protocol was followed for H&E staining of alternate sections (please refer to preparation of sections) on glass slides (object slides 26mm x 76mm x 1mm, Marienfeld, cat no.1000100). All steps were performed in room temperature:

1) Slides were mounted on a slide holder.

2) Slides incubated in Haematoxylin for 2.5 minutes.

3) Slides washed in distilled water for 45 seconds.

4) Slides incubated in acid alcohol for 45 seconds.

5) Repeat wash of slides in distilled water for 45 seconds.

6) Slides incubated in Scott's solution for 1.5 minutes.

7) Repeat wash of slides in distilled water for 45 seconds.

8) Slides were thereafter incubated in eosin for 2.5 minutes.

9) Wash of slides in distilled water for 45 seconds.

10) The holder is then immersed and incubated in the following freshly prepared solutions:

- 50% Ethanol for 1 minute.
- 70% Ethanol for 1 minute.
- 80% Ethanol for 1 minute.
- 90% Ethanol for 1 minute.

- 100% Ethanol for 1 minute.
- 2 sets of Xylene solutions for 1 minute each.

11) Slides were immediately mounted in DPX and covered by a coverslip. They were stored in a cool dry container for subsequent viewing.

1.2.10. Selection of appropriate sections for TIS runs:

Given the extremely heterogenous distribution of cancer cells in a tumour mass, we selected the process of detecting them by studying the immediately adjacent sections with protocol H&E staining. These H&E stained sections were studied by a trained histologist and presence or absence of cancer was reported in them. The approximate location of cancer cells in the H&E stained section was also pointed by a coloured "pointer" by the Histologist. Based on these findings, we selected the appropriate sections for TIS experiments. The "pointer" helped us in selecting the appropriate visual fields, as will be discussed below.

1.2.11. Preparation of tissue for experiments with TIS:

Following steps were followed in preparing the tissue section immediately before starting TIS run:

1) The sections were selected from the catalogue based on the results of H&E staining.

2) Slides with attached coverslips were taken out of the freezer.

3) These were incubated in ice cold acetone at minus 20^0 C for 10 minutes.

4) Following this the section was air dried for 2 minutes and clean tissue paper was used to wipe off the acetone around the section. Care should be taken not to disrupt the tissue section.

5) A silicone ring (prepared in house from "Press-To-Seal silicone sheets, adhesive on one side, Molecular Probes, cat no. P24725; rings with inner diameter: 10mm, outer diameter: 20mm and height: 1.4mm) was attached to the coverslip surrounding the

section, so that a watertight “well” was created and section lies at the bottom of it. The area of the glass coverslip around the tissue section should be wiped absolutely dry to facilitate the attachment of silicone ring.

6) The section was thereafter incubated with 200 μl PBS by filling the well with it for 10 minutes.

7) At this stage, the coverslip was carefully separated from the slide and the other side of the coverslip was wiped off all the embedding media by 70% ethanol solution. This was a critical step, as this surface would be in direct contact with the lens of the inverted microscope of TIS and in the light path of the microscope. Any excess embedding material on this surface can therefore result in the following errors:

- Artefacts in images.
- Impaired movement of the slide over the lens and restricted shifting of one visual field to another in between the run causing blurred images.

8) At this stage the PBS used in step 6 was removed and the tissue section incubated with 200 μl of goat serum solution (1.5 ml goat serum solution prepared by mixing 50 μl of goat serum + 1450 μl of PBS) for 30 minutes to block non-specific antigenic sites.

9) Following this the tissue section was washed with 100 μl PBS for 20 times.

10) At the end of the washing step, the well was filled with 200 μl PBS, rendering the cover slip ready for the TIS run.

1.2.12. Setting up of TIS run, preparation of antibody and buffer trays:

At this stage, an appropriate programme is written in which the exact tag names, incubation times, washing steps and the wells from which the robot collects the respective antibody/buffer solutions are specified in detail. According to the data in this programme, 96 well PCR plates (BrandTech, semi skirted 0.2 ml, 96 well PCR plate, cat. no. 781400) were used to prepare antibody solutions. Buffer (PBS) used for intermediate

washing steps was filled in the buffer trays (BrandTech 96 deep well plate, 2.2 ml, cat no. 701354), placed in TIS. Both these trays were covered with aluminium foil which would be pierced by robot disposable pipette tips (Eppendorf Combitips plus, 0.1ml, Biopur, cat no. 0030 069.404) in TIS. Following these steps, the coverslip with the prepared tissue section surrounded by the silicone ring is fixed (taped) to the platform over the inverted lens of the TIS microscope. Low power of magnification (x10 and x20) is used to select the appropriate area of the visual field initially. When an appropriate area in the section is selected in low power, it is focused against the high power oil immersion lens (x63). After manual selection of visual fields the TIS automated software is allowed to run repeated incubation-imaging-bleaching-cycles on the same set of visual fields (maximum of 5 visual fields), giving stacks of images as described before. The images are automatically saved in preselected folders in the hard drive of the TIS computer.

1.2.13. Construction of toponome maps:

As previously mentioned, for every antibody tag used (including each cycle of PBS), the system stored 4 images for each visual field concerned. They were named as IF (initial fluorescence), OF (post fluorescence), IP (Initial phase) and OP (post phase), as mentioned before. Every fluorescent image is then corrected for inhomogeneous illumination by using block based histogram-equalization algorithm, which is programmed in interactive data language. By processing the phase contrast images (IP and OP), relative misalignment in between images (due to shift between visual fields by movement of the microscope stage in the run) are corrected using a block matching algorithm with cross correlation as a maximum function. These corrections taken together results in a stack of “shifted” (aligned) and “subtracted” (net fluorescence) images. These images are then manually “binarised”, where an expert (chief researcher, assisted by a histo-pathologist) sets thresholds for each given fluorescent signal, revealing the most prominent binary signals, which are characteristic for any given

antibody tag, as discussed in details by Schubert and colleagues[373]. Superimposition of all these “shifted”, “subtracted” and “binarised” images for each visual fields by MoPPi software gives us a toponome map for the corresponding visual field. This map is presented as clusters of proteins and is viewed using the same software on the background of a phase image (to compare relative locations of different CMPs in the tissue) or over another fluorescent image (for example, DAPI, showing the relative positions of the nuclei and CMPs).

CHAPTER 3

1.3. RESULTS: OPTIMISATION OF TIS TO COLON CANCER RESEARCH.

Optimising first TIS unit at Warwick for colon cancer project:

In this section we will discuss the steps taken towards optimising the first TIS unit for a future project with colon cancer. We will also highlight the obstacles faced and technical errors for which several runs had to be abandoned. The aim of this section is therefore to help future researchers in this unit. The approval for this research has been granted on 26th October, 2006 by the Warwickshire Local Research Ethics Committee, Warwickshire, UK. The research has been performed in Biological Sciences Laboratory, University of Warwick, Coventry, UK. The human tissue has been collected from operative samples at the George Eliot Hospital, Nuneaton, UK.

1.3.1. Finalising antibodies and their adequate concentrations for TIS:

A pre-requisite for TIS, as already mentioned in "Methods" section, is antibodies primarily conjugated with FITC. Several antibodies like CD44, CEA, Cytokeratin-19, Cyclin A, Cyclin D1, PCNA, Bax, CD29, CD36 (from Santa Cruz Biotechnology Inc, Santa Cruz, USA), EpCAM, Muc-1, CD166 and Ki67 (from AbD Serotec, UK) were purchased as primarily conjugated form which had been previously used in human tissues and has been shown to work within a range of concentrations[410-412]. We started our optimisation steps with these antibodies on our trial colon cancer sections on the TIS using short 3 or 4 antibody programmes on one visual field. However, when we applied these concentrations to the TIS, we found that most of these antibodies failed to stain any tissue (Figure 8a, see below). Overnight staining of most of these antibodies in separate experiments on colon cancer tissue (sectioned by similar specifications)

revealed good staining. We implicated our failure to demonstrate adequate staining in TIS on the short incubation time (60 mins) as compared to overnight staining (10 hours). Longer incubation in TIS had the potential limitations of drying of tissue section being stained (progressively increased risk of drying towards the end of the run) and usage of fluorescence bulb time (maximum recommended length of use of one bulb = 300 hrs, HBO 100 fluorescent bulb, Zeiss, cat.no. 380018-4060, $160.00). These limitations were overcome by increasing the concentrations of these antibodies and finally reaching the

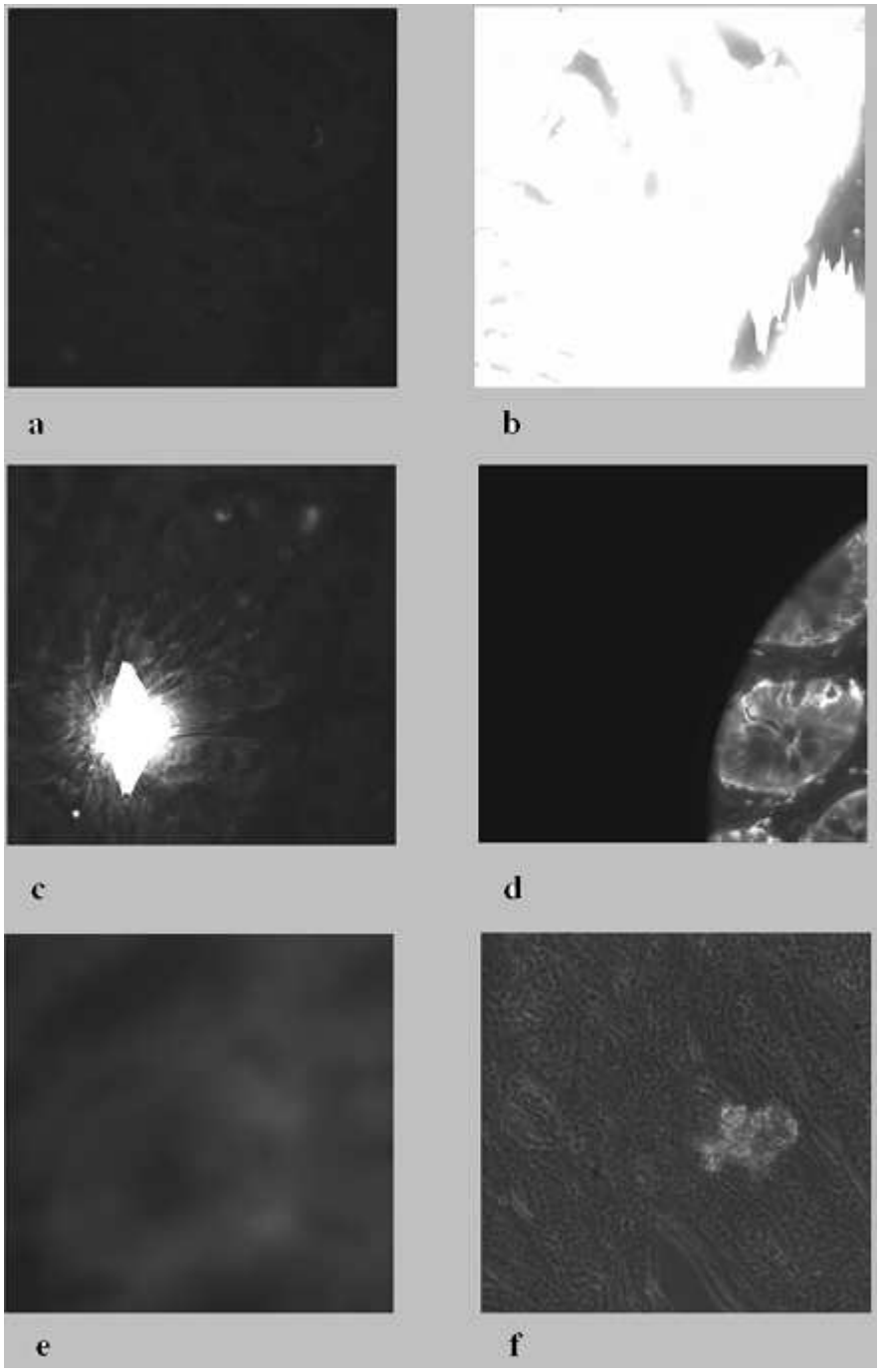


Figure 8: Figures showing results from optimisation runs. Figure 8a shows a section of colon cancer where antibodies did not localise any proteins due to low concentration during optimisation. Figure 8b shows complete white out image due to much higher concentration of antibody Ki67 used on a colon cancer specimen. Figure 8c shows an artefact with false fluorescence. Of note, these artefacts can result in false staining. Figure 8d demonstrates unopened shutter, a problem encountered during TIS runs in Warwick. Figure 8e shows a blurred phase image, possible causes being un co-ordinated shutter opening or mal-positioned condenser of the microscope or moving stage while imaging process in going on. Figure 8f shows artefact and blurring of phase image due to auto-focussing on the artefact rather than on the tissue section.

desired dilutions as mentioned in the “methods” section. On few occasions we also accidentally used much higher dilutions (especially with antibodies like CEA and Ki67) and these resulted in complete whiteout images (Figure 8b). The exposure time (camera shutter opening time) for fluorescence image was fixed at 1000 ms for all antibody tags (except DAPI, which was 100 ms). These fluorescent times were also titrated and optimised with different exposure times and antibody concentrations to reach these values to obtain the best images. Several antibodies like CD133, Cytokeratin-20, Musashi-1 (from Santa Cruz Biotechnology Inc, Santa Cruz, USA), Muc-2 (from AbD Serotec, UK) were not available in primary FITC conjugated form and had to be custom made (specially conjugated by the respective companies). Of these antibodies, Musashi-1 never worked in our tissues, a cause for which was never found. This tag was therefore left out from our subsequent experiments. Again, some antibodies although showed good staining initially, subsequent runs demonstrated huge artefacts (Figure 8f) possibly from impurities present in the antibody itself. As we had no facility or expertise of filtering antibodies in our unit, we had to dispose these antibodies as well.

1.3.2. Optimising TIS specifications:

The TIS programme that was responsible for running the automated repeated incubation-imaging-bleaching cycles had to be written before hand. The variables like shutter opening times (exposure time), light intensity, amount of PBS to be used in

washing steps, number of washing steps to be used in-between two cycles, temperature of the CCD camera and that of the TIS system were optimised one at a time depending on our past experiences and with help from our German collaborators (Dr. Schubert's laboratory in Magdeburg). The shutter opening times was the most significant of the lot as very short opening times (less than 100 ms) in a long run often resulted in one shutter remaining closed in several images, resulting in partial blackout of the corresponding segments of these images (Figure 8d). These short shutter opening times also resulted in un-coordinated opening of two shutters, which also resulted in extreme blurring of few images, rendering these images and therefore the whole run unusable. This problem was also resolved by setting the CCD camera temperature to minus 20^{o} C. The final exposure times as mentioned in the "Methods" section were 600 ms for all phase images (except that for DAPI which was 2000 ms) and 1000 ms for all fluorescent images (except 100 ms for DAPI). The light intensity for the microscope was optimised to 3.2 V. Intensities higher or lower than this value resulted in bright or dark phase images respectively. Poor illumination was responsible for errors in auto-focussing steps of the TIS microscope and resulted in blurred images (Figure 8e).

1.3.3. Optimising steps of tissue section preparation:

As already mentioned before, we have collected our colon tissue from (colectomy) operative specimen. Although this patient had full bowel preparation before the operation, in some cases faecal impurities were trapped. In some initial tissue sections prepared, we found these foreign bodies presenting as bright spots in all the images. However, by careful selection of visual field (avoiding solid particles in the visual field) in the sections studied, decreased the incidence of such foreign bodies (Figure 8f) in subsequent experiments. We believe that to reduce the incidence of foreign bodies in sections, colon tissue needs to be rinsed in distilled water carefully. To reduce artefacts in images resulting from impurities in PBS, we filtered the PBS (used for washing steps in the TIS run, diluting antibodies and goat serum and preparing tissue section before

TIS run) using cellulose filter paper (cat. no. CFP1-070, Sterlitech Corp. Kent, WA, USA). However, we still get some occasional artefacts in our images, the cause of which is currently being investigated.

1.3.4. Technical problems:

We encountered a few technical problems which are worth discussion. They were responsible for disruption of several experiments and causing major delays in this project. We started with a Micro Line ML-4710 camera with CCD sensor for imaging in TIS, supplied by EHD imaging GmbH (EHD imaging GmbH, Damme, Germany, originally supplied by Finger Lakes Instrumentation, Lima, New York, USA), which had 4 metallic shutter lamellas, fitted on a rubber ring. These shutters after several runs underwent corrosion and this along with dirt particles caused increased inter-shutter resistance causing un-coordinated opening of shutters on one or more occasions. The consequences of such happening have already been discussed in the section under "Optimising TIS specifications". This problem needed repeated cleaning of the shutters by a technical expert and abandoning the entire run. The technical support was un-available in our unit (funding issues). Hence the use of this camera was found to be unacceptable. This camera was replaced by a new camera (supplied by the same company), a high quality product from Uniblitz, USA, which had 2 lamellas. It performed much better, though extremely sensitive to the temperature. However when the temperature was fixed to minus 20^{o} C it worked better and posed no further problems with respect to image quality after the shutter opening times were optimised.

In this period, we also found that the TIS suddenly stopped producing any fluorescent images with the new CCD camera. This was found to have resulted from a cable fault which co-ordinated the fluorescent imaging and shutter opening. This needed further technical support from Germany.

The change of fluorescence bulb needed opening of the rear segment of the TIS cabinet and on one occasion was associated with displacement of the robot arm by few

centimetres. As a result of this, when we tried to start our next experiment, the TIS programme could not identify the robot and flashed **“Init COM4 error”** message. Due to lack of in house technical support, we had to wait for assistance from Germany and this was resolved by simple manual re-placement of robot arm to the correct position.

The **MoPPi software** is German in house software (from Magdeburg) and is slow to work on certain UK systems. It needs a bigger computer hard drive and does not work on laptops. This prohibits reproducing the TIS analysis in other computers and therefore restricts all the work on to one computer only. However, projects relating to its development are ongoing in Germany.

The TIS present with any errors as **“Init COM4 error”** and it needs technical expertise to know and solve the background problem, however small it may be. This causes major delays in the experiments. In house technical support may be an answer to this problem. An improvement in **“notification of the type of error”** (specifying exactly where the error is and what is the easiest way to resolve it) can also help in resolving it faster and resuming work on this system.

1.4. RESULTS: TIS analysis of colon cancer and normal colon from the same patient:

In this section, we demonstrate the potential of TIS in the investigation of colon cancer, a complex disease, involving disordered cellular biology which can be readily observed under the microscope. In this "proof of concept" section[413], a colon cancer tissue (5 μm section) was compared with another tissue section (5 μm) from adjacent area of normal colon, sourced from the same colectomy specimen, 10 cm away from the cancer edge. The tissues were respectively confirmed as cancer-containing and normal by an independent pathologist, experienced in colorectal cancer. These sections were prepared for TIS as previously described[373]. Serial sections were H&E stained and were also examined by the pathologist to confirm that cancer was present or absent in each tissue preparation (Figure 9x and 10x). Figure 9x, shows adenocarcinoma with infiltrating, irregular, glandular, neoplastic elements and forming abnormal acinar architectural structures. The carcinoma shows a moderately differentiated pattern, with constituent cells demonstrating abnormal cytological features, including disrupted cytoplasm with variable, enlarged, hyperchromatic nuclei showing some conspicuous nucleoli. In addition, nuclei are disordered and instead of being in a regular orderly basal arrangement are multilayered and stratified. Surrounding the glands a reactive fibro-inflammatory stroma is also present. Conversely, Figure 10x, shows an entirely normal colonic glandular structure, with no evidence of cancer and normal cellular architecture. The visual fields (VF) for this study were selected by detailed analyses of the histopathology in line with previous studies[379].

The various individual TIS images resulting from 21 different protein location cycles, following alignment and subtraction, are shown in Figures 9a - 9w and 10a – 10w. These images were then used for subsequent MOPPI clustering and other analyses to determine CMPs for normal and colon cancer tissues.

Cancer CMPs:

Following features could be noted in the list of CMPs obtained from this cancer run:

- The total number of CMPs in colon cancer was found to be 6813 (supplementary table).
- Most glandular epithelial cells in colon cancer were found to co-express Ki67 and CK-20, a feature found slightly more prominently in cancer as compared to normal (data not shown). However, when we looked at the co-expression of the cytokeratins (CK-19 and CK-20) in cancer (Figure 11), the motif showed marked decrease in cellular expression. The CMPs representing the cytokeratins were found to be grossly disorganized and clumped in the peri-nuclear area in cancer cells.
- Another feature of the colon cancer was a higher degree of expression of CEA and CK-20 combination CMPs on the epithelial cell surfaces as compared to normal (data not shown).
- In the cancer tissue, we found a motif of few busy CMPs, co-expressing CD133, CD44, EpCAM and CD166 as the lead protein. These CMPs highlighted five cells in the cancer tissue (Figures 12a and 12b, first row, marked by black arrows). Three (upper) of these cells were found to be adjacent to the glandular epithelial cells and other (lower) two, in the matrix. Interestingly, when CD133 was taken off this list of lead proteins, the resultant CMP motif (Figures 12c and 12d, middle row, marked by white arrows) highlighted only one more cell in the whole visual field in the matrix, apart from the above five cells (centre of the visual field, CMP colour = purple). This clearly shows that in most of the cells CD133 is co-expressed with the other three proteins in cancer.
- From the list of all CMPs in cancer (supplementary table), it has been found that the cells marked by all the above protein clusters also co-express Ki67 and most nuclear proliferative markers.

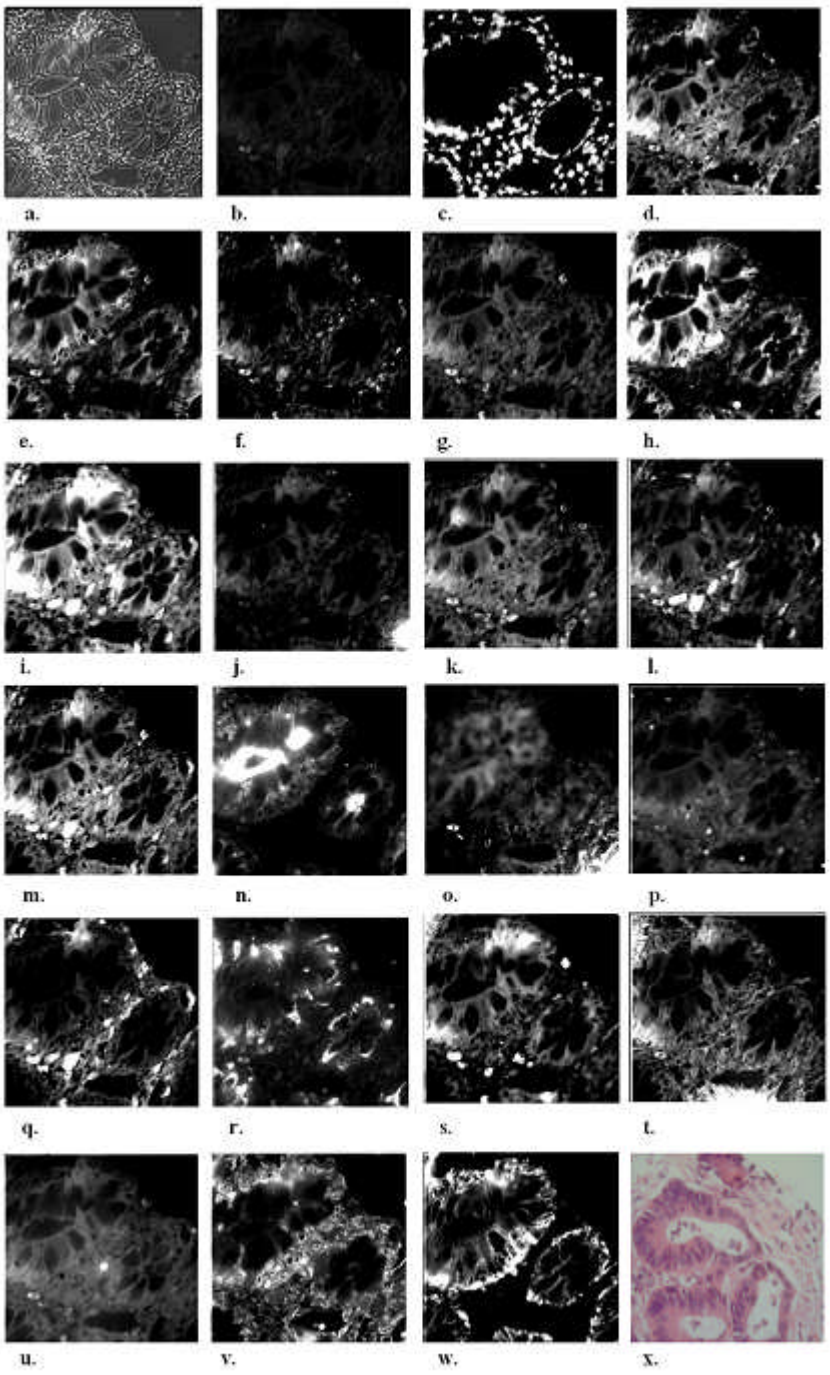


Figure 9: Fluorescent images from cancer colon section used to create cancer CMPs.

- We have studied the CMPs that are specifically expressed in the stromal cells (Table 3) and glandular cells in colon cancer (Table 4). We have obtained a motif of CMPs which highlight the cells present in stroma in particular with very little expression in the glandular epithelial cells. In this motif, the lead proteins were CD57, Laminin and

PCNA. The proteins CD29, CD36, CEA, CK-19, Muc2, Muc1 and NCAM were absent and hence “anti-co localised”. Rest of the proteins were variably expressed (wild cards). In contrast to stromal cells, the glandular cells mostly co-expressed EpCAM, CK-20 and CEA as lead proteins with anti-colocalisation of CD44 and CD36. The rest of the proteins were found to be variably expressed. These findings are shown in a schematic diagram in Figure 13.

CMPs in Normal:

Although, we primarily used normal colon to subtract the background normal features from colon cancer, we found that many features in normal colon were totally different from those in cancer tissue.

- The first striking feature was the number of total CMPs, found in normal colon was nearly 5 times (32,009, please see supplementary table) as compared to cancer tissue (6813, supplementary table).
- Among the 15 most frequent CMPs in normal colon, only a few CMPs were found to be very interesting. The CMPs 1, 3, 4, 7, 8, 9 and 11 (table in figure 14), had 17 or more different proteins co-expressed. NCAM and CEA were anti-colocated. Muc-2 and EpCAM were found to be wild card proteins in this CMP motif. When looked on the phase and DAPI images (Figure 14), interestingly most of these CMPs were localised in glandular epithelial cells. Very few of these busy CMPs were found to be expressed in the matrix. Another interesting finding was the total absence of these CMPs in colon cancer tissue.
- All CMPs with CK-19 and CK-20 as lead proteins express many more glandular cells in normal colon as compared to colon cancer (Figure 11c and 11d). When visualized on respective phase and DAPI image backgrounds the motifs with these lead proteins are extremely crowded on the glandular epithelial cell surfaces as compared to the cancer cells.

- In contrast to cancer tissue, only a few glandular epithelial cells showed CMPs co-expressing CEA and CK-20 (data not shown).

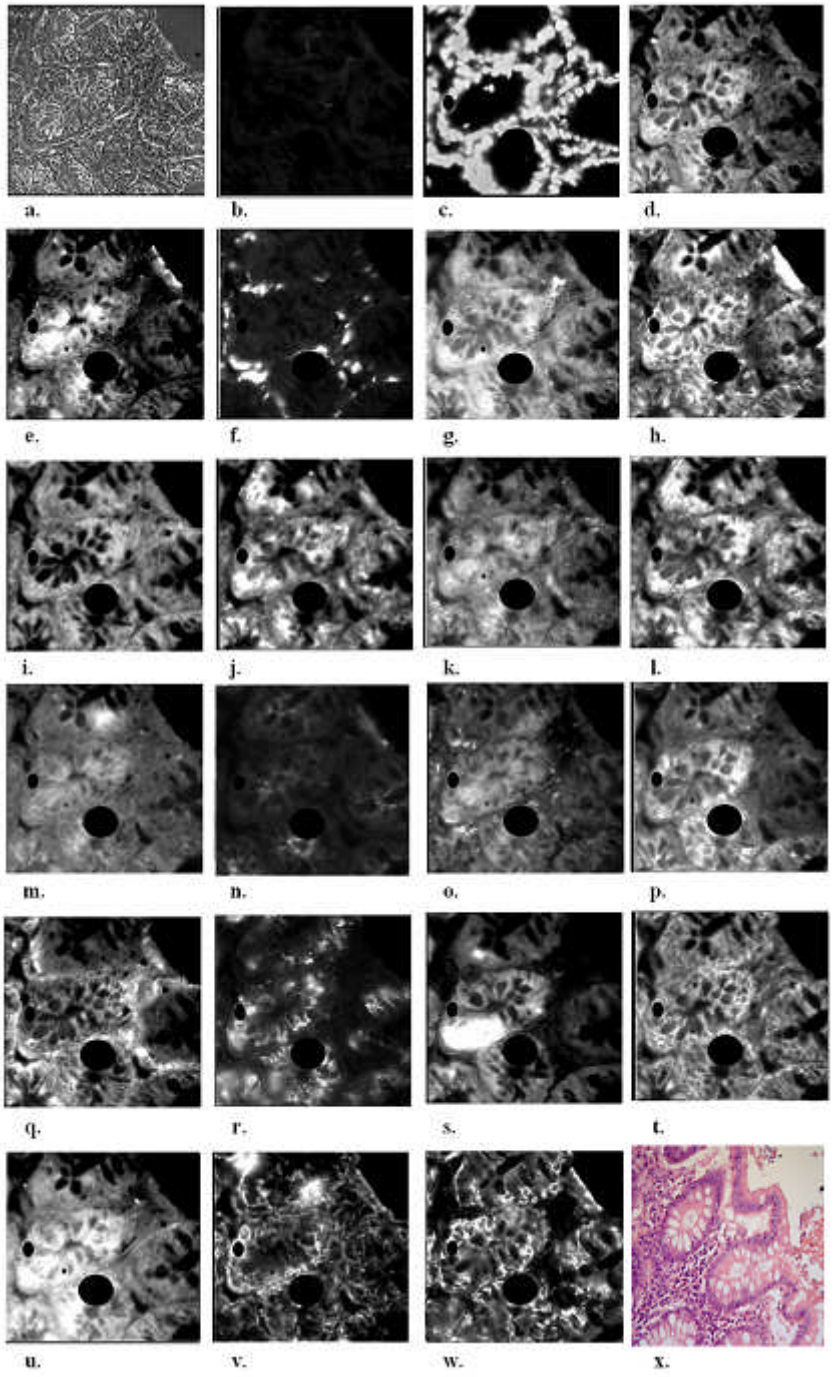


Figure 10: Fluorescent images from normal colon section used to create normal CMPs.

Cancer-unique CMPs:

By subtracting the CMPs found in normal from those in cancer (by Multicompare software of MoPPi), we found 5708 CMPs (full list in supplementary table), which were unique to cancer and will be referred as “cancer unique” motif. A list of 15 most frequently appearing CMPs in “cancer unique” motif is shown on the cancer tissue in figure 15.

In “cancer unique” data set, we found a motif of 34 CMPs (refer to Table 2 at the end of manuscript, Figure 16a and 16b, magnified in 16c and 16d on phase and DAPI backgrounds respectively). All these CMPs had the following lead proteins: CD133, CD166, EpCAM, Muc-1, Laminin, CD57 and CK-20. The common findings about these CMPs are as follows:

- Cells marked by these CMPs also expressed Ki-67, PCNA, Bax, Bcl-2 and Cyclin-A.
- CD44 was found to be a wild card protein and was seen to be co-existent with the rest of the lead proteins in only 5 CMPs (range of frequency 28 – 1).
- The cells marked by these 34 CMPs were mostly present adjacent to the glandular epithelial cells (table 5).
- All the CMPs described by those lead protein were found to be peri-nuclear or slightly overlapping the respective nuclei. This proves that the CMPs are marking specific cells and are unlikely to be spurious.
- A magnified view of these cells is given in figures 16c and 16d. In order to get a full list of proteins and CMPs present in these few cells of interest, we looked at them in the raw CMP data from cancer and found that, all these cells also co-express CD44, along with the above proteins. However, most of these CMPs with CD44 were found to be present in normal colon as well as cancer. Surprisingly only few CMPs (with very low frequency) in these cells had CEA, which was contrary to our expectations.

MoPPI analyses of Cancer and Normal CMPs:

The first level of analyses involves a comparison of the 15 most frequent CMPs unique to either cancerous or normal colon (Figures 15 and 14, respectively). In each of figures 14 and 15, the left panel shows CMPs superimposed on the phase contrast image; the right panel shows the same CMPs superimposed on the corresponding DAPI

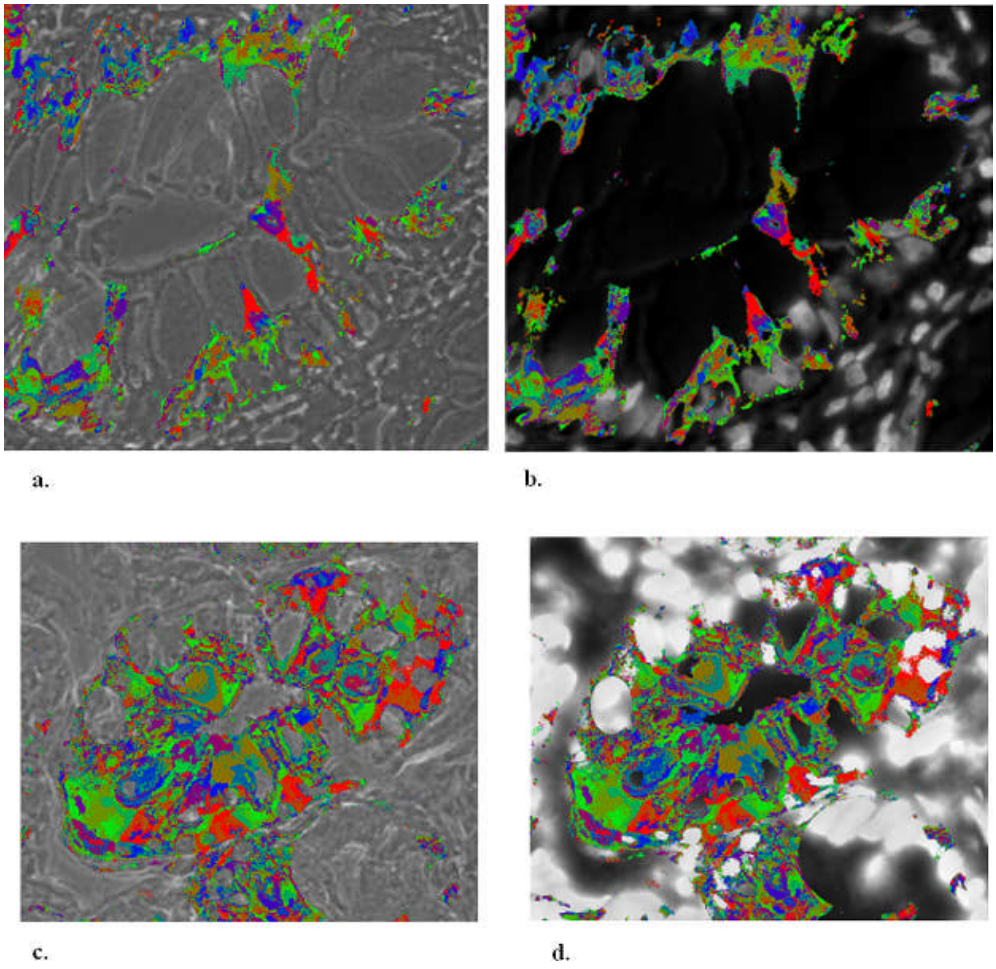


Figure 11: Glandular structure in: colon cancer (figure a, phase contrast and b, DAPI) and normal colon: (figure c, phase contrast and d, DAPI). The cell cytoskeleton, marked by CK-19 and CK-20, is disrupted in colon cancer (figures a and b) and clumped in the perinuclear area (figure b). In contrast, in normal colon the cell cytoskeleton is much more evenly distributed (figure c and d).

fluorescence image showing the location of cellular nuclei (white). CMP coding lists are shown on the bottom of the figures with (i) specification of molecules co-mapped by TIS (top line no.´s 1 – 21 of the list), (ii) specification of coloured CMPs, in which every CMP expresses a unique combinatorial binary code (1 bit information per protein/molecule)

corresponding to the colours in the phase- and DAPI- images; (iii) CMP motifs (bottom line of the list). For this study we chose to concentrate on the 15 highest frequency CMPs, based on previous studies in other systems.

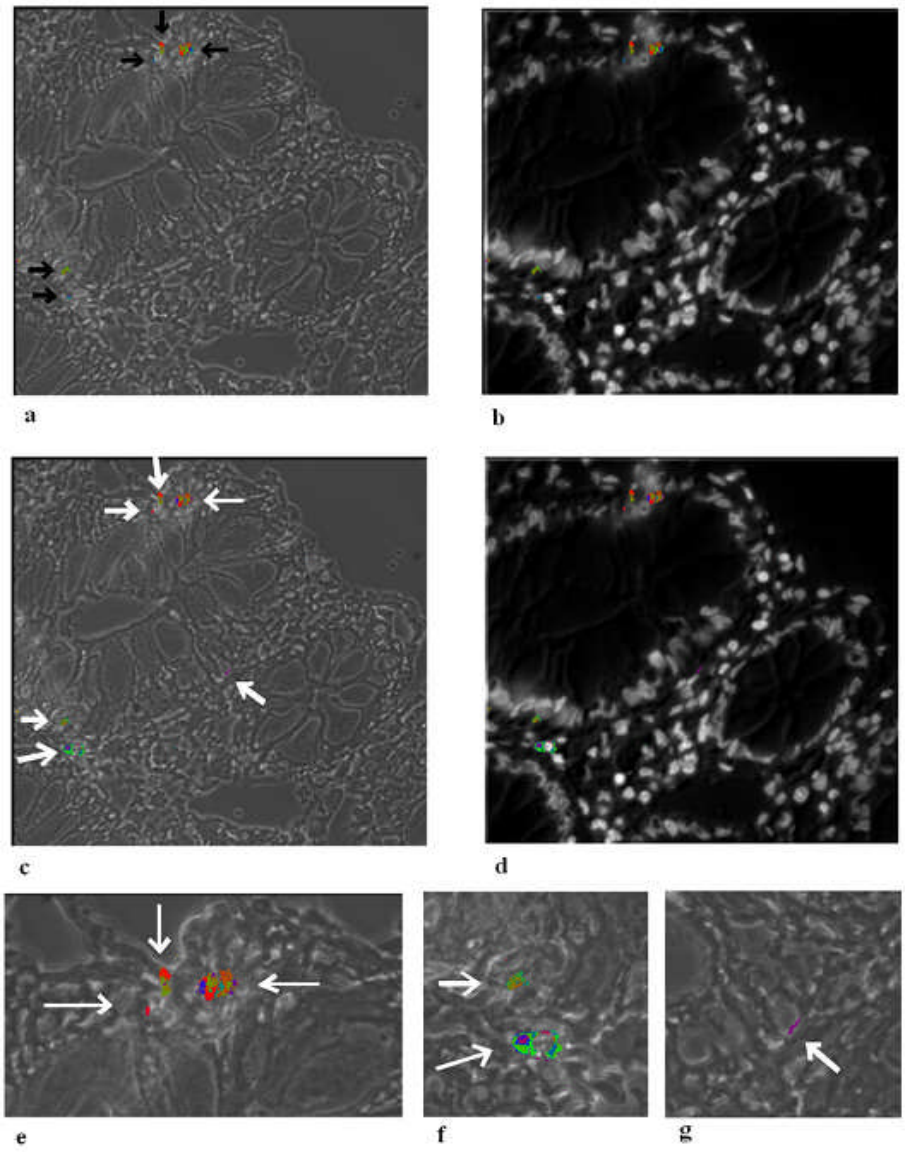


Figure 12: Subset of colon cancer cells identified by CMPs. Cells marked by CMPs with lead proteins; CD133, CD44, EpCAM and CD166 in the top row (a and b) and cells marked by all the CMPs with CD166, CD44 and EpCAM as lead proteins in the middle row (c and d) in cancer tissue. The CMPs in the top row have been shown by black arrows and those in the middle row by white arrows. The bottom row (figure e, f and g) shows magnified views of all CMPs from figure c. All the CMPs in the top and middle sections have CD133 co-expressed along with CD166, CD44 and EpCAM, except the single cell (CMP=colour purple) towards the centre of the middle row of images (c and d). This cell is magnified in figure g in the bottom row.

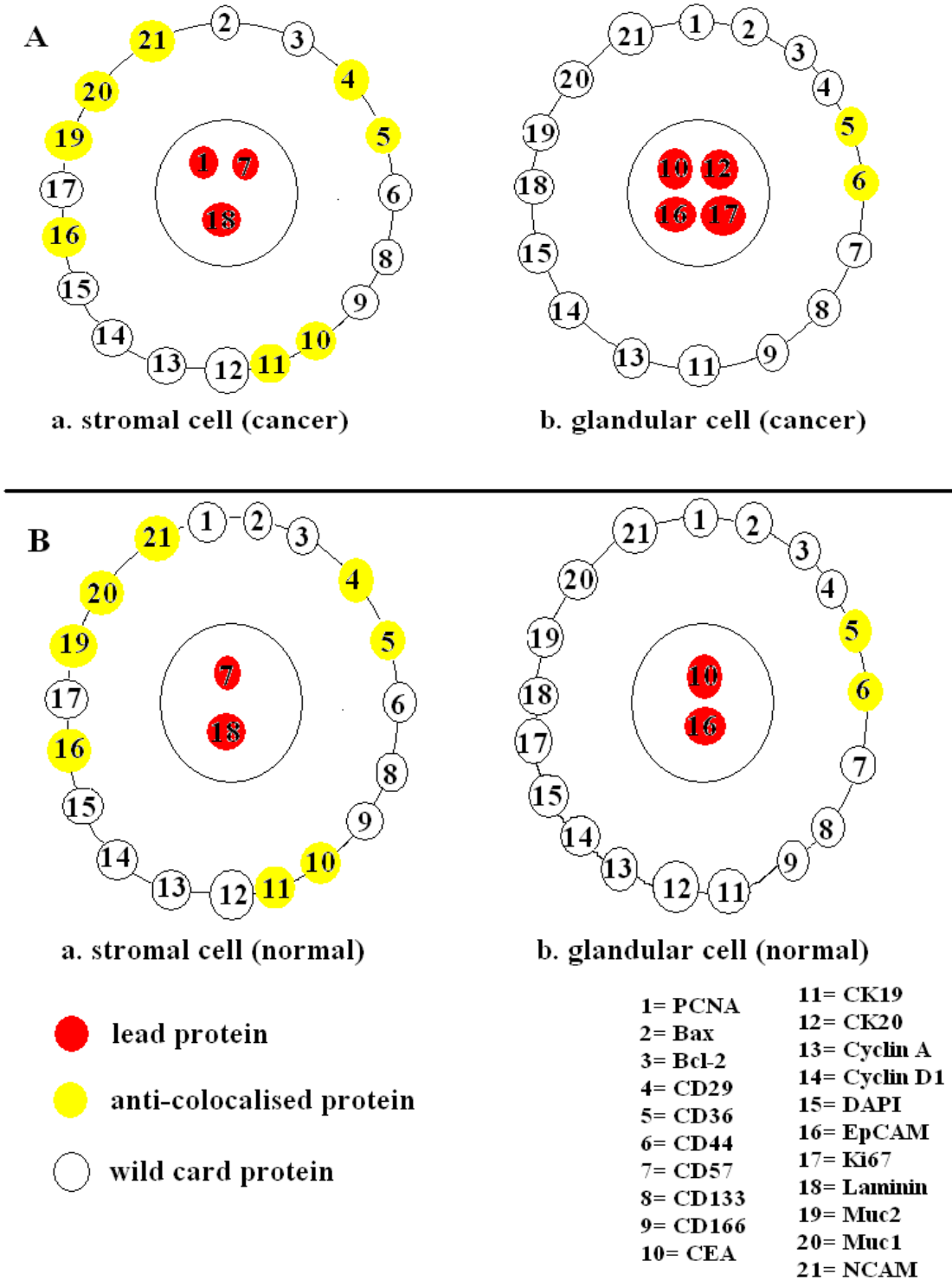
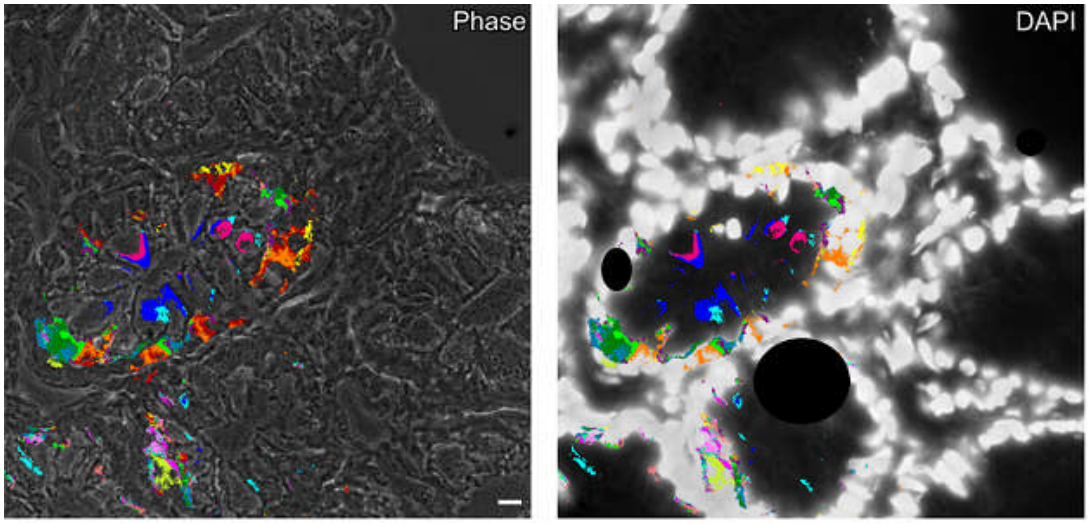


Figure 13: Schematic diagram indicating CMP motifs specific for glandular cells and stromal cells in colon cancer (in 13A) and normal colon (in 13B). The proteins are indicated by numbers as in the table in the bottom right hand corner. The inner circle shows proteins (highlighted in red) which are lead proteins. The outer circle shows the other proteins in the motif, the ones highlighted in yellow are anti-colocalised or absent and the rest are variably present (wild card proteins).

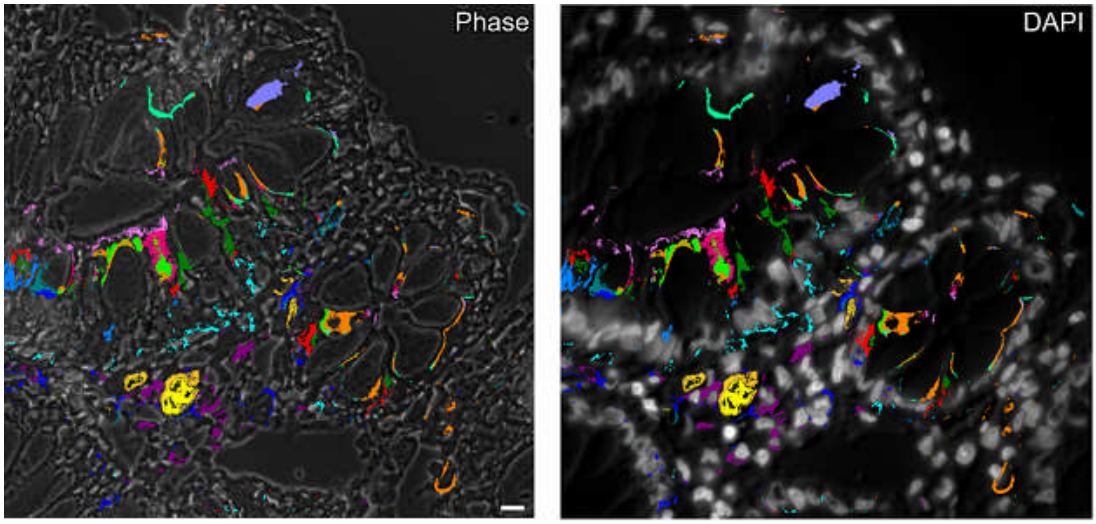


proteins / molecules

CMP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0
	1	0	0	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	1	0
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0
	1	0	1	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	0	0
	1	0	1	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	1	1	1
	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	0
	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	0
	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0
	1	1	1	0	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	0
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	0
	1	1	1	1	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	1	0
	1	0	1	0	0	0	1	0	1	0	1	1	1	0	0	1	0	1	0	1	0
	1	1	1	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	1	1	1
	1	1	1	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	1	0
Motif	L	W	W	W	W	W	L	W	L	A	L	L	L	W	W	W	W	L	W	W	W

1 PCNA
2 Bax
3 Bcl-2
4 CD29
5 CD36
6 CD44
7 CD57
8 CD133
9 CD166
10 CEA
11 CK19
12 CK20
13 Cyclin A
14 Cyclin D1
15 DAPI
16 EpCAM
17 Ki67
18 Laminin
19 Muc-2
20 Muc-1
21 NCAM

Figure 14: Fifteen most frequent CMPs in normal colon. In this figure, each colour represents one specific CMP with sub-cellular resolution. The left panel shows CMPs superimposed on the phase contrast image; the right panel shows the same CMPs superimposed on the corresponding DAPI fluorescence image showing the location of cellular nuclei (white). Note the CMP coding list shown on the bottom of this figure with (i) specification of molecules co-mapped by TIS (top line no.´s 1 – 21 of the list), (ii) colour specification of CMPs, in which every CMP expresses a unique combinatorial binary code (1 bit information per protein/molecule) corresponding to the colours in the phase- and DAPI- images; (iii) CMP motifs (bottom line of the list). This CMP motif represents the higher order organisation of the 15 different CMPs as a group with features common to all CMPs described by a three symbol code. **L**, lead protein/molecule (present in all CMPs); **A** (proteins/molecules which are absent in all CMPs = anti-colocated); **W** (wild card proteins/molecules variably associated with the L and A molecules of CMPs). Bar: 10 µm.



proteins / molecules

CMP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0
	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	1	1	1	0	0	0	1	0	1	0	1	1	0	0	0	1	1	1	0	1	0
	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	0	1	0
	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0
	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0
	1	1	1	0	0	0	1	1	0	0	0	1	0	0	0	1	1	1	0	0	0
	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
	1	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0
	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0
	1	0	1	0	0	1	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0
	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0
Motif	W	W	W	A	A	W	W	W	W	W	W	W	W	W	A	W	L	W	W	W	A

1 PCNA
2 Bax
3 Bcl-2
4 CD29
5 CD36
6 CD44
7 CD57
8 CD133
9 CD166
10 CEA
11 CK19
12 CK20
13 Cyclin A
14 Cyclin D1
15 DAPI
16 EpCAM
17 Ki67
18 Laminin
19 Muc-2
20 Muc-1
21 NCAM

Figure 15: Fifteen most frequent CMPs in colon cancer. In this figure, each colour represents one specific CMP with sub-cellular resolution. The left panel shows CMPs superimposed on the phase contrast image; the right panel shows the same CMPs superimposed on the corresponding DAPI fluorescence image showing the location of cellular nuclei (white). Note the CMP coding list shown on the bottom of this figure with (i) specification of molecules co-mapped by TIS (top line no.´s 1 – 21 of the list), (ii) colour specification of CMPs, in which every CMP expresses a unique combinatorial binary code (1 bit information per protein/molecule) corresponding to the colours in the phase- and DAPI- images; (iii) CMP motifs (bottom line of the list). This CMP motif represents the higher order organisation of the 15 different CMPs as a group with features common to all CMPs described by a three symbol code. **L**, lead protein/molecule (present in all CMPs); **A** (proteins/molecules which are absent in all CMPs = anti-colocated); **W** (wild card proteins/molecules variably associated with the L and A molecules of CMPs). Bar: 10 µm.

Illustration of the 15 most frequent CMPs which are unique to colon cancer-Figure 15 (not seen in the normal part of the same colon Figure 14; please see above):

Interestingly, the majority of these abnormal CMPs are located near to the apical part of the colonic epithelial cells and in the stroma of the colon tissue (in clear contrast to the normal-unique CMPs, which are primarily located at the basolateral structures of the colonic epithelium: see Figure 14). Note the extreme reduction of the number of molecules combined per CMP as compared to the normal colon (Figure 15 and 14, respectively). There is a cancer unique CMP motif with Ki 67 as lead protein, and a clear indication of a disruption of cytoskeleton/membrane molecule-expressing CMPs, as shown by the variably associated molecules 7-14 (W) in the motif, which is in striking contrast to the normal CMP motif (from Figure 14).

Illustration of the 15 most frequent CMPs which are unique to normal colon in Figure 14 (not seen in the cancerous part of the same colon in Figure 15):

In contrast to cancerous colon, the majority of the normal CMPs are located on the basolateral part of the colonic epithelial cells, whereas the cancer-unique CMPs, are largely located at the apical structures of the colonic epithelium (Figure 15). There is a normal - unique CMP motif with 7 lead proteins (PCNA/ CD57/ CD166/ CK19/ CK20/ CyclinA/ Laminin). The presence of these putative cytoskeleton/membrane molecule expressing CMPs, strongly supports the intact topology of cytoskeleton-membrane molecule assemblies. Note striking contrast to the cancerous part of the colon (Figure 15).

Total number of "Cancer-unique"CMPs:

When all CMPs found in normal were subtracted from all of those found in cancer using MoPPi, a total of 5708 cancer unique CMPs were identified and will be referred to as the "cancer unique" motif in this paper (The full list is given in supplementary table 1). The

15 most frequently appearing CMPs amongst the cancer unique group have already been shown in Figure 15. In the cancer unique data set, we found a motif of 34 CMPs (Shown in Figure 16 and tabulated in Table 2). All of these CMPs have the following lead proteins: CD133, CD166, EpCAM, Muc-1, Laminin, CD57 and CK-20. Cells marked by these CMPs also expressed Ki-67, PCNA, Bax, Bcl-2 and Cyclin-A, suggesting that these might be proliferative. The cells marked by these 34 CMPs were mostly located adjacent to the glandular epithelial cells. All the CMPs described by those lead proteins were found to be peri-nuclear or slightly overlapping the respective nuclei. CD44 was found to be a wild card protein and was seen to be co-existent with the rest of the lead proteins in only 5 CMPs (range of frequency 28 – 1).

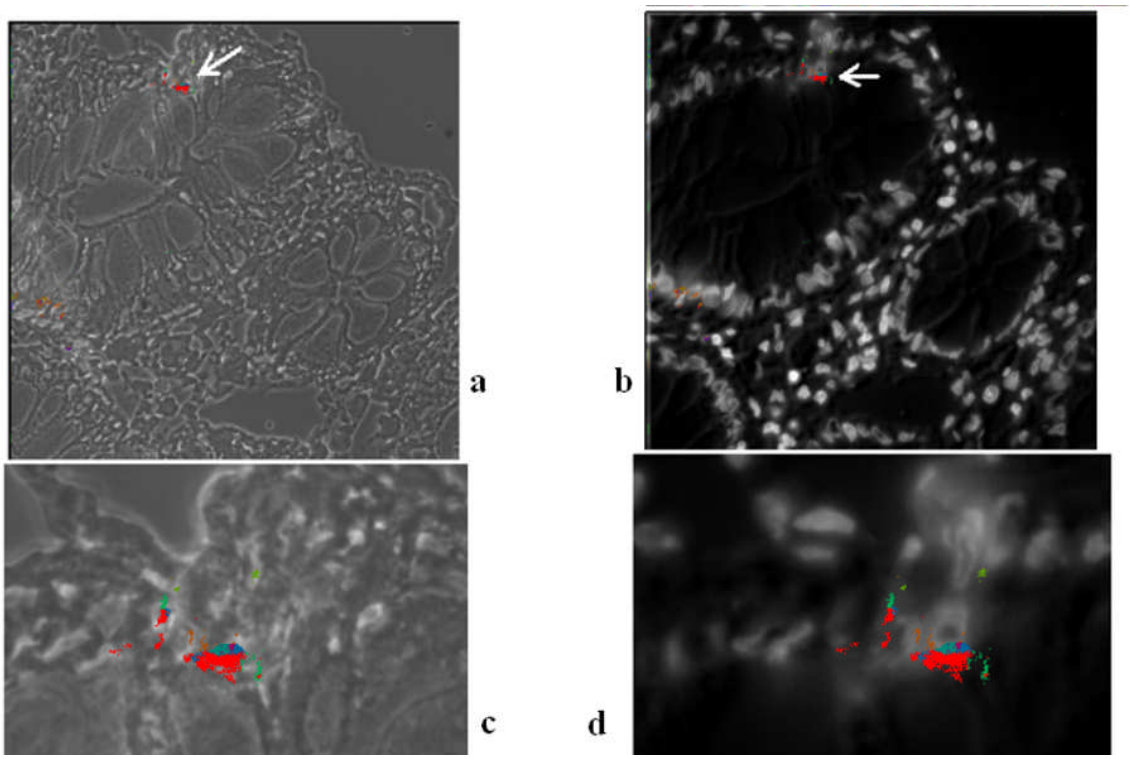


Figure 16 : **1st row;** All 34 Cancer unique CMPs in phase (a) and DAPI (b) images, with following lead proteins: CD133, CD166, EpCAM, CK-20 and Muc-1. Of note, these cells also co-express Ki67, PCNA, Bax, Bcl-2 and Cyclin-A. **2nd row;** CMPs in 1st row images are magnified in phase (c) and DAPI (d) images.

Normal unique CMPs:

Strikingly, the total number of CMPs, found in normal colon was nearly 5 times that (32,009) present in cancer tissue (6813), (supplementary tables 2 and 3, included in supplementary files only).The CMPs 1, 3, 4, 7, 8, 9 and 11 (table in figure 14), had 17 or more different proteins co-expressed.

CMPs may identify specific subsets of cells:

All the CMPs with CD133, CD44, EpCAM and CD166 as the lead protein highlighted five cells in the cancer tissue (Cells marked by black arrows in Figures 12, top row). The uppermost three of these cells were found to be adjacent to the glandular epithelial cells and the remainder were in the matrix. Removing CD133 from the CMP motifs (Figures 12c and 12d, middle row, marked by white arrows) results in the highlighting of only one more cell (CMP colour = purple), suggesting that in most of these cells CD133 is co-expressed with the other three proteins in cancer.

From the list of all CMPs in cancer (see supplementary table 3, in supplementary files only), it was found that the cells marked by all the above protein clusters also co-express Ki67.

Most glandular epithelial cells in colon cancer were found to co-express Ki67 and CK-20 (data not shown). However, when we looked at the co-expression of the cytokeratins (CK-19 and CK-20) in cancer (Figures 11a and 11b), the motif showed a marked decrease in cellular expression. The CMPs representing the cytokeratins were found to be grossly disorganized and clumped in the peri-nuclear area in cancer cells.

Many more glandular cells in normal colon as compared to colon cancer express CMPs with CK-19 and CK-20 as lead proteins. Moreover, when visualized on respective phase and DAPI image (Figures 11c and 11d) backgrounds, the motifs with these lead proteins are extremely crowded on the glandular epithelial cell surfaces as compared to the cancer cells.

In contrast to cancer tissue, only a few glandular epithelial cells showed CMPs co-expressing CEA and CK-20 (data not shown). We have also examined CMPs that are specifically expressed in stromal cells (Table 3 at the end of manuscript) and glandular cells in colon cancer (Table 4 at the end of manuscript). We have found a motif of CMPs which highlight the cells present in stroma in particular with very little expression in the glandular epithelial cells. In this motif, the lead proteins were CD57, Laminin and PCNA. The proteins CD29, CD36, CEA, CK-19, Muc2, Muc1 and NCAM were absent and hence anti-co localised. Rest of the proteins were variably expressed (wild cards). In contrast to stromal cells, the glandular cells mostly co-expressed EpCAM, CK-20 and CEA as lead proteins with anti-co-localisation of CD44 and CD36. The rest of the proteins were found to be variably expressed. We have also shown the expression of these same proteins in normal colon in the form of three symbol code (Figure 13B). Please note that in contrast to cancer tissue, proteins Ki67 and CK20 are variably expressed (wild cards) in glandular cell motif in normal colon. The anti-colocalised proteins are the same in both cases. These findings are shown in a schematic diagram in Figure 13 above.

CHAPTER 4

1.5. GENERAL DISCUSSION AND FUTURE PROSPECTS

Colorectal cancer is characterised by uncontrolled growth of colorectal epithelial cells. According to the theory of multistep carcinogenesis, colorectal epithelial cells accumulate a number of molecular changes and eventually become fully malignant cells through steps that has been discussed above. Genetic and epigenetic events during the carcinogenesis process differ considerably from tumour to tumour. Thus, colorectal cancer may be considered to be a conglomeration of heterogenous complex of diseases with different sets of genetic and epigenetic alterations. Essentially, each tumour arises and behaves in a unique fashion that is unlikely to be exactly recapitulated by any other tumour. Therefore level of expression of different proteins in colon cancer varies widely and has been extensively studied by different authors. Below we have discussed some of these proteins which we have used in our study with Toponome Imaging System.

1.5.1. Colon cancer and related proteins:

CD133 (prominin-1) is a trans-membrane glycoprotein (120kDa, 865 amino acids), encoded by a gene on chromosome 4 in humans (4p15.33). It has an N-terminal extracellular domain and a intracellular segment composed of 59 amino acids, the two connected by five trans-membrane domains with two extracellular loops[414,415]. It is thought to play a role in organizing and maintaining the lipid composition of plasma membrane of intestinal epithelial cells[416]. The significance of CD133 in the field of colon cancer stem cell can broadly be in three fields.

- Its potential role in identifying stem cells from a group of several colon cancer cells, which may be a step towards discovering a new stem cell bio-marker.
- Its functional significance in colon cancer, which is currently under much debate.

- Its potential prognostic significance in colorectal cancer patients.

Reports from Marzesco and colleagues suggested that expression of CD133 in a cell sustains its stem cell phenotype. As CD133 containing vesicles are released, the cell becomes committed and starts differentiating[417]. Yin and colleagues produced a monoclonal antibody, AC133, which identified a specific glycosylation-dependent epitope of CD133[418]. Though, many human cell lines and differentiated cells express CD133, its AC133 epitope is only expressed in stem or progenitor cells and not in mature cells as the glycosylation status of a cell changes on maturation or malignant transformation. Thus AC133 epitope may be a more specific marker for stem cells[419]. According to the stem cell theory, cancers originate from normal stem cells or from cells, which are their immediate descendants[420]. Independent reports published by two research groups have confirmed the presence of such cancer initiating cells in human colon expressing CD133[37,38]. Subcutaneous injection of immuno-deficient mice with freshly isolated, $CD133^+$, human colon cancer cells readily reproduced the original tumour from which the cells were extracted. Interestingly, serial transplantation of these $CD133^+$ cells produced tumours which had similar phenotype but were much more aggressive than their parent tumours. Floating undifferentiated $CD133^+$ cell aggregates called colon spheres resulted when colon cancer cells were cultivated in a proliferative serum-free medium. Upon differentiation, these cells expressed high levels of colon cancer specific marker, CDX2 and cytokeratin-20 (CK-20, a protein expressed in differentiated gastric and intestinal epithelial cells) along with progressive down-regulation of CD133 expression. Similar attempts with $CD133^-$ cells failed to induce tumorigenesis in immuno-deficient mice. These experiments indicated that only undifferentiated $CD133^+$ cells (colon cancer stem cells) are capable of tumorigenesis[37]. Reports from *in vivo* limiting dilution experiments by O'Brien and colleagues indicated that on an average one cancer stem cell exists in every 262 $CD133^+$ cells in a colorectal tumour and in every 5.7×10^4 un-fractionated colon cancer cells[38]. Interestingly, these cells retaining tumorigenicity *in vitro* were resistant to many chemotherapeutic

drugs[37,421]. However, conflicting results have been published which shows that CD133 is not only exclusive to colon cancer stem cells but also expressed in other colon cancer cells (unable to replicate in xenograft host) along with a range of normal adult human tissues like colonocytes, pancreatic ducts, proximal tubules in kidneys and lactiferous ducts of mammary glands[40,422].

Nikolova *et al* reported that proliferation of cord blood derived $CD133^+$ cells are exaggerated in culture media containing Wnt signalling factors (Wnt 1, Wnt 3a, Wnt 4, Wnt 5a and Wnt 11) and is associated with increased expression of neural stem cell marker nestin and endothelial cell surface markers (CD31 and von Willebrand factor). Furthermore, cells grown in Wnt 3a containing media demonstrated a more undifferentiated phenotype. This paper also reported that Wnt 5a induced expressions of nuclear and cytoplasmic β-catenin m-RNA and protein[423]. This evidence supports a direct link between Wnt /β-catenin pathway and proliferation in cells expressing CD133. Similar reports have established links between Hedgehog, Notch and BMP signalling pathways with CD133 expressing cancer stem cells in human brain tumours[424-426]. The significance of these morphogenetic signalling pathways in colon cancer has already been discussed. These evidences point towards a functional basis for the expression of CD133 in colon cancer stem cell.

Increased expression of CD133 in colon cancer tissue correlates well with low patient survival [427,428]. This has been supported by a recent publication showing increased incidences of liver metastases in patients with high expression of CD133, both in the primary cancer and the corresponding liver secondary. By gene knock-out studies the authors demonstrated that CD133 has got no significant functional contribution to the aggressiveness of colon cancer cells *in vitro*[429]. However, they failed to demonstrate the expression of CD133 in normal colon, which differs from those of a previous report[40].

Cell adhesion molecules (CAMs) are known to play significant roles in cell cycle, cell migration, cell signalling, tissue regeneration and interaction between cell and the

matrix. Thus they maintain tissue homeostasis. They are also involved in the pathogenesis of cancers, inflammations and auto-immune diseases. The CAMs are divided into 4 super-families: Cadherins, Selectins, Integrins and Immunoglobulin-CAM (IgCAM). Apart from these four super-families there are also some other CAMs, which do not share the characters of any specific family (eg. EpCAM)[430]. The first example from CAMs playing active role in cancer metastasis came from the family of CD44. Invasive cancer phenotype is often associated with loss of E-Cadherin[431].

The cloning of CD166/ALCAM was described for the first time in 1995 [(432)]. CD166 is a trans-membrane glycoprotein (65kDa, 583 amino acids) and a member of IgCAM superfamily. It has an intracellular cytoplasmic chain (of variable length), a transmembrane domain and an extracellular component consisting of 2 distal variable and 3 proximal constant Ig domains. The main functions of this protein are ligand binding, cellular avidity and development. Ligand binding is related to the extracellular domains and it is responsible for both homo and heterophilic cell-cell adhesion (ALCAM-ALCAM and ALCAM-CD6 respectively). Developmental studies have revealed that CD166 is mostly expressed in proliferating cells. Its normal tissue distribution in selected epithelial cells, lymphoid and myeloid cells, bone marrow mesenchymal cells, hepatocytes, pancreatic islets and acinar cells, cells of para-aortic mesoderm in developing embryo supports this function. This protein has also been expressed in several cancers like colorectal, breast, lung, prostrate and melanomas. Though its function is somewhat similar to cadherins in terms of cell adhesion, interestingly, in contrast to the later group, CD166 is found to be present and often induced in several cancers, especially colorectal. This degree of induction may also co-relate with the invasiveness of the respective cancers concerned[433]. Weichert *et al* established connection between membranous expression of CD166 and shortened survival in colonic adenocarcinoma[434].

CD326/EpCAM is a type 1, glycoprotein (40kDa). Like other CAMs, it has extra cellular, trans-membrane and intra cellular domains (ICD). Its distribution in fertilized egg,

morula, embryonic stem cells and cells undergoing differentiation with concomitant absence in terminally differentiated cells supports its role in embryogenesis and differentiation. It promotes cellular proliferation, which is dependent on minimal growth factors and in an anchorage independent manner. In keeping with this property, it is over-expressed in carcinomas (except malignancies of mesodermal or ectodermal origin). Its over-expression in pancreatic and colorectal cancer is associated with poor prognosis and invasive/metastatic disease [435]. It is a homotypic adhesion molecule and antagonizes cadherin mediated cell-cell adhesion [436]. The ICD fragment of EpCAM has been implicated in well recognized cell membrane to nucleus signalling pathways, however exact mechanisms are still un-clear. Munz *et al* demonstrated that inducible expression of EpCAM can provoke expression of *C-MYC* and its target genes like *Cyclin E* and *Cyclin A*[437]. Of note, products of both these genes are well recognized proliferating factors. Therefore, induction of *c-myc* and its target genes may explain the observation that, EpCAM expressing cells demonstrate rapid proliferation in an anchorage independent manner and with a reduced requirement for growth factors[437,438]. Maetzel and colleagues have demonstrated that ADAM17 (metalloprotease) and γ–secretase sequentially cleave EpCAM to produce its 5kDa ICD fragment. From these *in vitro* experiments, the authors reported the presence of EpCAM–ICD fragment in cytoplasm and nucleus of the experimental proliferating cells, which could be prevented by using specific protease inhibitors in cell culture. They also found the existence of these EpCAM–ICDs in nuclei of colon cancer cells in contrast to normal colonic cells[439]. In a short review published in 2009 by Carpenter and Brewer[440], EpCAM has been aptly referred to as a "Surface to Nucleus Missile". The mechanism by which EpCAM–ICD induces proliferation of cells is thought to involve protein FLH-2 (four and half LIM domain protein 2). Reports confirm the presence of EpCAM–ICD as an intranuclear (650 kDa) complex, which also includes FLH2, β-Catenin and Left1/TCF. Exogenous addition of EpCAM–ICD without EpCAM, also results in formation of this intranuclear complex

with induction of *c-myc* and its target genes[441]. As discussed under the section of β-Catenin, Left1/TCF is a major regulator of *c-myc* expression. Based on these findings, related to the significance of EpCAM–ICD and cleavage of membranous EpCAM by proteases like (ADAM17 and γ–secretase), EpCAM is being considered as a therapeutic target in several cancers and currently several anti EpCAM agents are being used in clinical trials[442].

CD44/HCAM is a trans-membrane glycoprotein and is expressed as several iso-forms, of which the commonest is the standard or haematopoietic form. All the isoforms belonging to this family are encoded by a single gene on chromosome 11 in humans and chromosome 2 in mice. CD44 isoform proteins are mostly expressed in cells of neuro-ectodermal origin and are characterised by a constant region comprised of large extracellular domain composed of 270 amino acids, trans-membrane region formed of 23 amino acids and an intra-cellular domain with 72 amino acids. These regions are encoded by the first and the last 5 exons and account for the smallest but most commonly expressed isoform, CD44s. A variable region, close to the transmembrane region, is encoded by various combinations of exons 6–15 (v1–v10), giving rise to CD44 variant isoforms (CD44v). The N-terminal of the ectodomain in all isoforms is characterised by a hyaluronic acid binding site. The ectodomain is extensively modified by N- and O-glycosylations, which affect the HA binding ability of CD44 isoforms[443]. Function of the HA binding domain in these protein molecules is dependent on the cells concerned, for example, hyaluronic acid turnover by mesenchymal cells and “homing” of leukocytes during inflammation[444]. The ectodomain of CD44v3 (an isoform of CD44) is decorated with heparin sulphate, which enables it to bind growth factors like VEGF[445]. The intracellular domain (ICD) of CD44 has a binding site for Ankyrin. and Ezrin-Radixin-Moesin (ERM) proteins. Both of these interactions provide a link between CD44 and the cell cytoskeleton. The ankyrin binding site is close to the domain of amino acids that anchors ERM proteins to CD44. Phosphorylation of a serine residue at position 291,

close to the ERM binding site, inhibits binding of ICD to ERM and thus influences CD44-mediated directional cell motility[446,447]. CD44 proteins are known to play active role in cell proliferation, differentiation, survival and motility. A CD44 protein isoform, containing the exons v4–v7 was identified in a highly metastasizing rat pancreatic carcinoma cell line, BSp73ASML. Interestingly, these specific variants of CD44, when transfected into related BSp73AS cells with no metastatic potential, were found to confer the metastatic potential to the later upon injection into syngeneic rats. Furthermore, CD44 exon v6-specific antibody also blocked the metastasis in these cells. Animals treated with metastatic cells BSpASv4–v7 with the CD44v6 antibody blocked lymph node and lung metastases in them. This was the first major finding pointing towards an active role of CD44 isoforms in cancer metastasis[448,449]. In human colon, expression of CD44s as well as other variants is reported to be enhanced in adenomas and carcinomas. Expression of CD44s, CD44v3, CD44v6 and CD44v8–v10 also correlates with bad prognosis and advanced disease in patients[450]. In an *in vivo* murine breast cancer model, the HA binding capacity of CD44 ectodomain was shown to promote metastasis and inhibit apoptosis. CD44 – HA molecular aggregates were essential for binding MMP9 (Matrix metalloprotease 9) to the cancer cell surface and induce the activity of MMP9. This in turn, induced collagen IV degradation and subsequent invasion of these cancer cells. Activated MMP9 on cancer cells also activated TGFβ leading to angiogenesis of cancer tissue[451,452]. In an *in vitro* model, invasion of colon carcinoma cells was also found to be dependent on CD44 binding to HA and on accumulation of HA in the matrix in peri-cellular region. Invasion of these cells in matrigel was prevented by CD44 antibodies, preventing CD44 – HA interaction and also by introducing an antisense construct against 2 of the 3 human enzymes (HAS2 and 3) responsible for HA synthesis[453]. CD44v6 and CD44v5 are up-regulated in gastric cancer and CD44v5 expression is preferentially found in poorly differentiated carcinomas and metastatic lymph nodes. Presence of CD44v3, v6, v7/8 in breast cancer correlates with more aggressive stages of the disease[454,455]. Interestingly, CD44 has also been shown to interact with collagen,

laminin, fibrinogen, chondoitin sulphate, fibronectin, serglycin / gp600, MHC Class II (major histocompatibility complex class II) invariant chain, selectins (N-Selectin and E-Selectin) and osteopontin (OPN). Osteopontin is a cytokine and is found to contribute to metastasis in various cancers. OPN specifically binds to CD44v6, which leads to enhanced cell motility and chemotaxis. In in vitro experimental model of hepatocellular carcinoma, CD44 blocking antibody inhibited invasion of cancer cells, which was induced by osteopontin[456-458]. CD44v6 has been tried as a target for anti-cancer therapy in squamous cell carcinoma (SCC) in humans, where frequent and homogeneous expression of this isoform was detected. A humanised version of an anti-CD44v6 antibody labelled with rhenium-186 was tested in patients suffering from head and neck squamous cell carcinoma (HNSCC). Based upon the initial promising results, bivatuzumab coupled with a non-radioactive cytotoxic drug, bivatuzumab mertansine, was developed. Thirty patients were treated with this compound. This trial demonstrated disease stabilisation and regression of tumour in 10% cases. Although an unfortunate mortality in a patient in this cohort due to toxic epidermal necrolysis terminated this project prematurely, CD44v6 and other isoforms still continue to be molecules of potential interest for therapeutic modulation of different cancers[431,459,460].

Cytoskeletal filaments are formed of Intermediate filament (IF) proteins, acting as their building blocks. Depending on the amino acid sequence and structure of protein, IF proteins are subdivided into 5 major types. Types I–IV IFs are cytoplasmic in location. Cytokeratins (CK) are the largest subgroup of IF proteins and are included under type I and type II or the acidic and basic CKs, respectively. They are functionally important in maintaining normal epithelial architecture and physiology in cells. Desmin in muscle tissue and Vimentin in mesenchymal cells are examples of type III IF. Type IV IFs include neurofilament proteins in neuronal cells. Type V IF, the lamins, are nuclear in location in contrast to the rest of the IFs. Several features distinguish cytokeratins from other IF proteins[461,462]. First, cytokeratins are preferentially expressed in epithelial cells; and second, they are obligate non-covalent hetero-polymers that include at least one of

each type I and type II keratin[462,463]. However, the cytokeratins possess a similar structural organization as compared to all other IF proteins. This includes a central domain formed of coiled alpha-helix, flanked by non–alpha-helical head and tail domains. These later non-alpha-helical domains contain most of the motifs that are responsible for regulatory modifications like glycosylation and phosphorylation, resulting in structural heterogeneity of the CKs[462,464]. The CKs which are predominantly expressed in simple (single-layered) epithelial cells or simple epithelial keratins (SEKs) consist of CK7, CK8, CK18, CK19, CK20, and CK23. Of these, types 7, 8, 18, 19 and 20 are expressed in normal gastrointestinal tract epithelial cells[465]. Simple epithelial cells are commonly found in glands and organs involved in secretion and absorption. These individual cells are often polarized, which suggests that the unique expression of SEKs in these cells is likely to have significant functional significance in these cells[466,467]. Among the cytokeratins, the most studied SEKs are CK8 and CK18[468]. Variable levels of cytokeratin 7, 19 and 20 are also expressed in many simple epithelial cells. Adult hepatocytes and proximal renal tubular cells express cytokeratin 8 and 18 exclusively, in contrast to enterocytes, which manifest a complex distribution that depending upon the stage of differentiation and specific cell type[469,470]. More recently identified, CK23 is found to be expressed in pancreatic and colonic tumours[471-473]. CK8 and CK18 are found in pseudo-stratified and urothelial epithelium, mostly in luminal cells. CK7 and CK19 serve as markers for ductal epithelial cells and are also found in gastrointestinal epithelium. CK20 is expressed primarily in terminally differentiated enterocytes, gastric epithelium, urothelial cells and neuroendocrine cells[469]. Under-expression of all four isoforms of CK-19 and over expression of CK-20 has previously been shown to be a feature of colon cancer[474]. Some cell types exhibit a polarized distribution of SEKs within distinct sub cellular compartments. For example, cytokeratins 8 and 18 localize preferentially within cytoplasmic filaments, whereas CK7 and CK19 are primarily found in the apicolateral compartment in close proximity to the cell membrane[475].Interestingly, CK19 and CK20 were reported to be up regulated in pancreatic injury and incorporated

into cytoplasmic filaments. This localization reverts to the typical apicolateral distribution upon healing[476]. Stress conditions and different pathological changes can affect not only keratin expression profiles, but also their expression levels and post translational modification. For example, increased keratin phosphorylation is a marker of tissue injury and disease progression in intestinal goblet cells in mouse and hepatocytes in human and mouse models[470,477]. Multiple SEKs are generally expressed in an epithelial cell–specific manner. This and the fact that cytokeratin expression profiles (but not their levels) remain relatively stable even during neoplastic transformation explain why cytokeratins are commonly used tumour markers. For distinguishing between tumours of epithelial and non-epithelial cell origin, staining with a pan keratin–specific antibody has been used widely. Metastases of specific tumours express certain specific cytokeratins. Secondary hepato-cellular carcinomas typically express the variants 8 and 18, and cholangio carcinomas express CK7 in addition to these above two cytokeratins. In contrast, metastases of colorectal carcinomas are mostly found to express CK20, but not CK7[469,478,479]. The expression pattern of different cytokeratins can also be useful in predicting the prognosis in certain tumours. In colorectal carcinoma for instance, low CK20 expression is encountered in carcinomas with high microsatellite instability, which constitute a distinct 15% entity of these cancers and are associated with improved survival compared with carcinomas with low microsatellite instability[480]. In breast cancer, expression of cytokeratin 8 and 18 is associated with favourable prognosis, whereas in breast cancers with CK5, CK14, and K17 the prognosis is relatively poor[469,478]. In the normal bladder, expression of CK20 is confined to the superficial umbrella cells, and also in non invasive transitional cell carcinoma. The persistence of this expression pattern predicts that the tumour is unlikely to recur[469]. Several reports suggest a direct beneficial impact of CK18 expression in breast cancers. For example, when this specific variant of cytokeratin is experimentally down regulated in advanced and high-grade tumours, its suppression directly correlates with shorter patient survival[481]. Similarly, *in vitro* studies revealed that CK18 over expression in a breast cancer cell line led to

reduced invasiveness of cancer cells and reduced tumorigenicity in nude mice models *in vivo*[482]. Although additional evidence is needed in this field, it appears that SEK expression levels may positively correlate with lower tumour invasiveness and SEKs may be a potential therapeutic target for modulating cancers in near future[483].

Mucins are high molecular weight glycoproteins, produced by epithelial cells. Mucin molecules include an oligosaccharide motif attached to the threonine or serine residue of its core protein backbone by *O*-glycosidic linkages. Mucins are classified into the following[484,485]:

- Membrane-associated mucins (MUC1, MUC3, MUC4, MUC12, MUC16, and MUC17),
- Gel forming secreted mucins (MUC2, MUC5AC, MUC5B, and MUC6), and
- Soluble secreted mucin (MUC7).

The epithelial membrane bound mucins are quite distinct from the classic extracellular complex mucins secreted by goblet cells forming the mucous layers of the gastrointestinal and respiratory tracts. These membrane associated mucins are in reality, trans-membrane (TM) molecules, expressed by most of the glandular and ductal epithelial cells[486]. They are characterised by an extensive extracellular domain, single trans-membrane segment and a cytoplasmic tail. The extracellular portion of these mucin molecules are composed of variable number of the tandem-repeat domain (TR domain), sperm-protein, enterokinase and agrin domain (SEA domain) and epidermal growth factor like domain (EGF-like domain). These TR domains are unique to mucins and differentiate the mucins from all other membrane bound glycoprotein molecules. The specific biochemical and biophysical properties of mucin molecules depends on the extent and nature of glycosylation of the TR domains[484,485,487]. Mucins have been studied extensively in relation to human cancers. Mucins have been reported to have complex associations with various cellular pathways related to proliferation, and apoptosis[488]. Transformation of fibroblasts and influencing tumorigenicity and metastatic potentials in

various cancer cell lines by experimentally modulating up regulation and suppression of MUC1 and 4 *in vitro* established this link in various scientific reports[489-493]. MUC1 undergoes aberrant glycosylation, is over expressed in 90% cases of breast cancer and is frequently expressed in other human cancers like colorectal, pancreatic, hepato-biliary, lung and ovarian[494]. In 1993 invasive ductal carcinoma of pancreas was reported to express MUC1 (pan-epithelial type membrane-associated mucin) in contrast to MUC2 (intestinal type secreted mucin), the later being predominantly expressed in intraductal papillary mucinous pancreatic neoplasm with indolent biological behaviour[495]. This and several other subsequent reports from other human solid tumours (mainly pancreatic and hepato-biliary) led to the conclusion that expressions of MUC1 and MUC4 independently correlated with more invasive forms of tumours with worse prognosis, in sharp contrast to MUC2, which was mainly found in tumours which were primarily non-invasive and had a better outcome[496-506]. MUC4 expression has been found to carry a poor prognosis for lung adenocarcinoma[507]. MUC16, also known as CA-125 is used as a tumour marker for ovarian cancer. It is the largest membrane bound mucin[508]. In a patient comparison study of mucin expression profiles, in between invasive intra-hepatic cholangio-carcinoma (ICC) with poor prognosis and non invasive mucin producing bile duct tumour (MPBDT) with a favourable outcome, MUC1 was predominantly found at the cell apex, lateral cell membrane, cytoplasm, and in luminal contents of the ICC cells. On long term follow up, these cancers demonstrated aggressive lymphatic permeation and peri-neural invasion. In sharp contrast to these findings, MUC2 was located in the cytoplasm of MPBDT cells only. All these cancers were characterised by less aggressive lymphatic metastases and no peri-neural invasion[509]. A similar study studying mucin expression in 136 gastric cancer patients, reported that MUC1 and MUC2 expressions were independently related to worse and favourable outcomes respectively[500]. MUC1 has been reported to serve as a natural ligand for galectin-3 in human cancer cells. Interaction in between cancer cell associated MUC1 and circulating galectin-3 enhances adhesion between endothelial surface and cancer cells, thereby promoting

metastasis[510].The cytoplasmic tail of MUC1 consists of 69 amino acids and includes several serine, tyrosine and threonine phosphorylation sites which are potential regions for interaction and activation of proteins associated with proliferation and suppression of apoptosis[508,511,512]. Previously we have discussed the significance of MEK / ERK kinases and β-catenin in the morphogenetic pathways associated in colon cancer. Through interacting with and subsequently activating ErbB1 growth factor, MUC1 activates ERK 1 and 2[513,514]. MUC1 has also been found to regulate ERK signalling by modulating the transcription of genes encoding for MEK-1, c-jun and Raf-1, by an ErbB-independent pathway[515]. Interaction of MUC1 with FGFR3 (Fibroblast growth factor receptor 3) induces tyrosin phosphorylation of the cytoplasmic tail of MUC1. This in turn, enhances binding between MUC1 and β-catenin and the transportation of the later in the nucleus. MUC1 also inhibits GSK3β-mediated phosphorylation and degradation of β-catenin and thus increases its cytoplasmic and nuclear levels. As previously stated, nuclear β-catenin acts as a transcriptional co-activator and increases the expression of genes related to proliferation like, *cyclin-D1* and *c-myc*[516,517]. MUC1 promotes cell proliferation in breast cancer cell lines (*in vitro*) by interacting with and stabilizing estrogen receptor α (ERα) and also stimulating ERα mediated transcription of growth factors and apoptosis inhibitors. This interaction and stabilization is induced by estrogen[515,518]. Recent reports suggest that MUC1 confers protection against oxidative stress induced cell death. This it does by regulating p53-responsive genes. As previously discussed, activation of p53 target genes promotes induction of apoptosis or a growth arrest and repair response. MUC1 selectively down regulates transcription of genes responsible for apoptosis by directly binding to the p53 regulatory domain[519]. In addition, MUC1 also induces anti-apoptotic Bcl-XL and PI3K/Akt pathways[520]. FGF1 induces tyrosine phosphorylation of the cytoplasmic tail of MUC1 and facilitates the binding of MUC1 to the heat shock protein (HSP) 90, finally targeting the cytoplasmic tail fragment of MUC1 to the mitochondria[516]. This interaction between MUC1 cytoplasmic tail and mitochondria is

associated with the attenuation of stress-induced release of mitochondrial pro-apoptotic factors, activation of caspase-3, and finally leading to cell death[521]. This suggests that by interacting with HSP90, MUC1-cytoplasmic tail transduces signals from the cell membrane to mitochondria, activating the intrinsic apoptotic pathway in response to stress[516]. Activated cytoplasmic c-Abl binds to cytosolic 14-3-3 proteins and translocates into the nucleus, where it activates *MEK kinase-1* and thereby activates the pro-apoptotic c-Jun N-terminal kinase (JNK) pathway[522,523]. Raina and colleagues have shown that MUC1 blocks this nuclear transfer of c-Abl and thus the apoptotic response to genotoxic anticancer agents is blocked[524]. Ahmad *et al* reported that MUC1 interacts with the IκB kinase complex, increasing the phosphorylation and degradation of IκBα, thus promoting the activation of nuclear factor-κB (NF- κB)[525]. The role of pro-carcinogenic role of NF- κB has already been discussed before. MUC1 also binds directly to caspase-8 and FADD, blocks recruitment of caspase-8 to the death-inducing signaling complex and thereby directly prevents activation of the extrinsic apoptotic pathway[526]. The functional significance of this interaction in normal epithelial cells is a protective mechanism, inhibiting unnecessary caspase-8 activation. However, malignant cells exploit this phenomenon for survival under adverse conditions by over-expressing MUC1[527].

1.5.2. TIS analysis of single section of colon cancer and its comparison to corresponding normal colon.

In the “proof of concept section” we aimed to confirm, for the first time, that the novel TIS platform could be successfully employed to investigate the architecture of a complex protein system in colon cancer *in situ*. Our first specific goal was to confirm that unique CMP patterns could be identified using TIS that correlate with and may illustrate certain key cellular biological features, as has been previously suggested for prostate cancer[379]. Ultimately, we hope that this preliminary study will act as a foundation for a full scale toponome study of colon cancer. The protocols, techniques and initial observations

presented here can readily be exploited now in larger scale validation and functional studies *in vivo* and *in vitro*. Here we have developed and optimised the TIS platform and then have presented the results of comparing a colon cancer tissue section with a matched control normal tissue taken from the same patient (some 10 cm away from the cancer), in order to minimise any inter- individual variation in protein expression. Absence of cancer in the control sample was confirmed by examining of H&E stained sections and further confirmed by a pathologist experienced in colorectal cancer. Following TIS imaging, protein co-localisation patterns (Combinatorial molecular phenotypes- CMPs) were identified in both cancer and normal sections and then compared by proprietary image-analysis software (MoPPi). CMPs are groups of pixels or voxels, signifying the localization of given protein clusters in a tissue section[195,373]. It is important to note that unlike regions of expression of a single protein epitope, CMPs represent areas of individual cells or structures, in which more than 1 protein is expressed, in very close proximity (within 1-2 pixels) to one another. Although, this does not prove that such proteins are physically interacting, the repeated finding of CMPs comprising particular proteins, when combined with anatomical observations of the cells/sub cellular compartments involved, can allow such a hypothesis to be generated. This provides a further new way for unravelling functional protein networks in systems biology approaches, because unlike most existing 'Omics' techniques, anatomical and spatial data are not lost as a consequence of the analysis. In a given combination of CMPs, proteins which exist or are entirely absent in all the clusters are termed as "lead proteins" (L) and "anti-colocalised" (A) proteins respectively. The rest of the proteins are variably expressed (present in some and absent in others) and are termed as "wild card" (W) proteins. A combination of CMPs thus can have one or more lead proteins or anti colocalised proteins and several wild card proteins, depending upon the number of protein tags used in the experiment. Such a combination of CMPs expressing a common feature is a "CMP motif". Most importantly, the grouping of CMPs into motifs with lead proteins reveals, for the first time, the functional structure of protein systems *in situ,* in

which the interaction of proteins within and between different CMPs is hierarchically controlled by the lead protein(s), as demonstrated for Rhabdomyosarcoma cells[371,372]. Blockade of a cell surface lead protein leads to disassembly of the corresponding network of interlocked protein clusters and loss of function (transition of the spherical cell to the exploratory state preceding migration / metastasis, APOCON model).

One of the key findings in this study has been the identification of multiple high frequency unique CMPs for both 'cancer' and 'normal'. The top 15 unique CMPs for each are shown in figures 15 and 14 respectively. Importantly, these CMPs correlate not only with normal and cancer but also seem to illustrate cellular biological differences apparent visually between the conditions, namely the different localisation sub-cellularly of the CMPs and also the link with cytoskeletal disruption in the cancer (see also Figure 11). This lends weight to the application of methodologies like TIS to pathological states where protein expression changes can be powerfully linked to histo-pathological /cytological behaviours/changes visible under light microscopy. This is inherently very different to the situation pertaining to many 'omics' techniques where disruption/destruction of the tissue of origin precludes any such simultaneous combined analyses of anatomy and molecular phenotype.

Our data clearly shows that in cancer, as compared to normal colon, the cell cytoskeleton system (marked by combination of CK19 and CK20) is grossly disrupted (Figure 11, see above in results section). Intermediate filaments are seen mostly clumped around the respective nuclei. In comparison to cancer, normal glandular cells show a more evenly distributed cell cytoskeleton system. Though cells in both the cancer and normal tissue expressed CEA (mostly at the apical cell surfaces) and CK-20, we found a higher number of CMP motifs containing both CEA and CK-20 in the glandular epithelium of colon cancer as compared to normal colon. This is important, as our results are consistent with those of other researchers[408,528]. As discussed above, cytokeratins are intermediate filaments and are important in maintaining normal epithelial architecture and physiology in cells. Few specific cytokeratins (types 7, 8, 18,

19 and 20) are expressed in normal gastrointestinal tract epithelial cells[465]. Under-expression of all four isoforms of CK-19 and over expression of CK-20 has previously been shown in colon cancer[474], and supports our own observations that CK20-CEA combination motifs are more and CK19-CK20 less frequent in colon cancer CMPs. We have also identified motifs that appear to distinguish neoplastic glandular and stromal cells (Figure 13, see above in results section). In this figure, the proteins indicated by numbers highlighted by red in the centre are lead proteins, those in the outer circle highlighted by yellow are absent from the CMPs and the rest are variably expressed (expressed in some and absent in other CMPs or wild card proteins).

Overall, 5708 unique CMPs were identified in cancer from patient A (complete CMP list is shown in supplementary table), and which were absent in normal colon. The degree of combinatorics per data point was found to be much higher for normal colon from the same patient as compared to cancer. Thus, a far greater number of CMPs were found in normal and this is exemplified by seven busy CMPs from the most frequent in normal colon (CMPs 1, 3, 4, 7, 8, 9 and 11 shown in the table in figure 14), which show that nearly 17 different proteins were co-expressed in different combinations in all of them. However, as these CMPs are largely localised to normal glandular epithelial cells and are totally absent from cancer glandular epithelium, it is possible that these CMPs are in some way inhibited by the process of carcinogenesis.

Although, still contentious, a large body of evidence suggests that colon cancer may be driven by a side population of putative colon cancer stem cells (CCSC). CCSC represent a small number of cells within the cancer, but are held to be essential for proliferation and maintaining the cancer mass[529,530]. In the last decade, numerous potential protein markers for CCSCs have been described. However, despite progress, none of these markers, individually or in low level combinations, have proved to be cancer-specific and are found, though maybe with reduced frequency, also in normal colon [38-39]. Thus, to date, no unique cancer stem cell protein signature has been described. However, some

very important predictions can be made, based on the wealth of important existing studies in this area. Thus, CCSCs are predicted to:

- Be located adjacent to epithelial cells.
- Express, amongst others, combinations of CD133, CD166, EpCAM, and CD44 on the cell surface.
- Express replication markers.
- Express additional markers essential for their unambiguous recognition and characterisation.

With this in mind, in this study, we have identified a small subset of colon cancer cells (five cells in total in the entire visual field), by use of specific CMP motifs with CD133, CD44, EpCAM and CD166 as lead proteins (figures 12a and 12b, please refer to results section). A further single cell was identified by presence of CD44, EpCAM and CD166 as lead proteins without CD133 (figure 12c and 12d, results section). Given, that these cells also express replication markers and are completely absent from the normal colon, these findings are entirely consistent with the predictions about putative CCSCs. There are some interesting points to emphasise here and which strengthen the likely relevance of the results, in particular the confirmation of previous observations from the literature. In brief, of all cancer unique CMPs: 2317 CMPs contained CD133, 1103 contained CD133 plus CD166 and 262 contained CD133 with CD44 as lead proteins and only 100, with all 3. Moreover, a higher degree of correlation exists between expressions of CD133 together with CD166 as compared to CD44. Interestingly, much of this is consistent with the existing literature, which suggests that CD133 is not always co-expressed with other stem cell markers[39], and that expression of both CD133 and CD166 is poorly correlated with expression of CD44 in colon cancer patients[531]. Many of the cells identified by stem cell marker protein containing CMPs in this study, also carry CMPs containing replication markers such as Ki67, PCNA, and Cyclin A, suggesting that they might be actively proliferating cells. This is also consistent with findings of previous

studies[37,39,529,532]. Our current findings also support the view that the search for stem cells may best be undertaken by employing a battery of markers, because single markers (for example CD133)[2, 40] or even a low level combination of markers[39]. However, lack of functional studies makes it difficult to confirm the same. These intriguing findings are subject of further studies by our team at Warwick. First, we will have to test the hypotheses generated here in a large cohort of colon cancer patients and second we will aim to exploit cell surface TIS signatures described here to isolate such cells by FACS or other allied techniques for functional study in future.

Our findings in this study also revealed that if a search for stem cells is performed by a single marker (for example CD133) or even by a group of few markers, this is likely to locate many nonspecific cells both in cancer and normal (data not shown). This finding supports the results of Shmelkov *et al*, who clearly demonstrated that CD133 is expressed not only by cancer stem cells but also by other cancer cells not possessing any stem cell property[40].

TIS analysis of colon cancer is a potential way to unearth huge data in stem cell research. Though lack of functional studies is a potential drawback with TIS, this system allows us to investigate intact tissue sections and identify the presence or absence of thousands of protein clusters in a single cell, implications of which are numerous. There are several future possibilities, aside from application of increasing numbers of different antibody libraries to allow ever more detailed molecular phenotyping of colon cancer cells. By combining the unique strengths of both TIS and of mass spectrometry based proteomics will be a very powerful one. The strength of TIS is the higher resolution, the ability to examine combinatorial protein expression at the cell surface and also the complete preservation of anatomical information. The strength of MS-based techniques, amongst others, lies in the ability to discover relevant new or previously unsuspected proteins in a given system.

Many more sections are being studied and extensive Toponomic database is being created. This is of extreme significance as any future work in this line can be compared

against this data set. Although, most of the experiments performed in last 2 years failed due to reasons of problems associated with imaging (in at least one or more visual fields of the entire data set), malfunctioning antibodies supplied by companies (may be related to the conditions to which they were subject while transfer from these companies to our lab), artefacts in tissue sections and custom made antibodies (discussed in optimisation section), extensive work is under way at Warwick to overcome these hurdles. A probable solution may be to create an in house system of FITC conjugation and generating antibodies for these experiments. Metastatic colon cancer in other sources (example omentum, liver, lymph nodes, lung, etc) can also be analysed by TIS. Colocalisation of similar motifs in cancer and metastases in omentum or other secondary sites from same patients and their absence in respective normal colon tissue may further strengthen the stem cell theory of cancer origin. This project however would need further ethics committee approval and funding. Colon cancer has been shown to arise in colon affected by ulcerative colitis[533]. Therefore projects comparing colon with ulcerative colitis and cancer can contribute significant information to the scientific community in this respect.

We have the potential to increase our antibody library. Our project involved optimisation of several antibodies used for the first time in TIS for human colon. Most of these antibodies were custom made for this purpose. The resulting cost implication was a potential limitation to the size of antibody library in our current project. This can be addressed by entering into industrial collaborations with companies to increase the supply of primarily conjugated antibodies. A relatively inexpensive alternative exists in setting up an in house conjugation unit as mentioned above, which can substantially reduce the expenditure and can ensure a ready supply of antibodies. However, this involves funding and employing full time technicians for this pupose.

Project involving software modification is currently underway in Magdeburg, where MoPPi software is being improved to work faster and automatically generate toponome maps as soon as the TIS runs are over. Automated binarising of net fluorescence

images is also a potential project in this field. An important hindrance in this aspect is the luminal micro particles and foreign bodies in colon which often result in false fluorescent signals and needs to be excluded from the final data sets. A similar in house project at Warwick University is currently under way. In this project the data images undergo registration and segmentation, followed by which the DAPI images are primarily used to generate the data. The aim of this project is to assess the TIS results with improved software which will provide quantitative information on the proteins expressed (shown as a graph) based on the level of fluorescence signals from the respective tags. However, this method will only be applicable when the antibodies and their concentrations used are fixed for every run being compared and thus needs further optimisation for respective tissue sections.

In-vivo and in-vitro experiments, functional studies: TIS project in cancer research can be taken a step further by introducing *in-vivo* and *in-vitro* experiments along with. The clusters thought to be significant in sections of cancer or diseased tissue (for example CMPs signifying stem cell subset) can be blocked *in-vivo* or *in-vitro*. The results from such blocking action can be further assessed by TIS analysis of cells from cultures (*in-vitro*) or biopsy material from animal models (*in-vivo*). It goes without saying that these results have serious implications in future therapeutic modulation.

To conclude, TIS is a novel technique which can reveal several important facts about the cellular biology of colon cancer and potentially also about different subsets of cancer cells, including potentially the elusive CCSC. It may help reveal much about protein interactions, identification of lead proteins, and most importantly TIS is a technique that can optically resolve the structure of cancer specific protein systems with direct implications for the hierarchical functions/dysfunctions of proteins. A fixed section can withstand a TIS run for at least 2 weeks and hence experiments using vast numbers of antibody tags are technically possible and would provide us with tens of thousands of CMPs[372]. These findings provide the rationale for a colon cancer toponome project on a proteome wide scale.

2. Published data and Ethics approval

A major portion of this thesis has been published in 2010 as a scientific publication by our group in Journal of Proteome Research. This publication is cited in Pubmed as follows:

Toponome imaging system: in situ protein network mapping in normal and cancerous colon from the same patient reveals more than five-thousand cancer specific protein clusters and their subcellular annotation by using a three symbol code.
Bhattacharya S, Mathew G, Ruban E, Epstein DB, Krusche A, Hillert R, Schubert W, Khan M.J Proteome Res. 2010 Dec 3;9(12):6112-25. Epub 2010 Oct 28.PMID:20822185

A portion of the introduction section on relationship in between fatty acid metabolism and colon cancer has also been published as an independent review article by our group and cited in pubmed as follows:

15-lipoxygenase-1 in colorectal cancer: a review.
Bhattacharya S, Mathew G, Jayne DG, Pelengaris S, Khan M.
Tumour Biol. 2009;30(4):185-99. Epub 2009 Sep 10. Review.PMID:19752603

Ethics committee approval for research

The approval for this research has been granted by the Warwickshire Local Research Ethics Committee, Warwickshire, UK. The research has been performed in Biological Sciences Laboratory, University of Warwick, Coventry, UK. The human tissue has been collected from operative samples at the George Eliot Hospital, Nuneaton, UK.

3. Bibliography

[1] Bhattacharya, S., Vijayasekar, C., Worlding, J., & Mathew, G., Octreotide in chemotherapy induced diarrhoea in colorectal cancer: a review article. *Acta Gastroenterol Belg* 72 (3), 289-295 (2009).

[2] Ricci-Vitiani, L., Pagliuca, A., Palio, E., Zeuner, A., & De Maria, R., Colon cancer stem cells. *Gut* 57 (4), 538-548 (2008).

[3] Voutsadakis, I.A., Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2. *J Cell Mol Med* 11 (2), 252-285 (2007).

[4] Adams, J.M. & Strasser, A., Is tumor growth sustained by rare cancer stem cells or dominant clones? *Cancer Res* 68 (11), 4018-4021 (2008).

[5] Horvitz, H.R. & Herskowitz, I., Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* 68 (2), 237-255 (1992).

[6] Watt, F.M. & Hogan, B.L., Out of Eden: stem cells and their niches. *Science* 287 (5457), 1427-1430 (2000).

[7] Moore, K.A. & Lemischka, I.R., Stem cells and their niches. *Science* 311 (5769), 1880-1885 (2006).

[8] Blanpain, C., Horsley, V., & Fuchs, E., Epithelial stem cells: turning over new leaves. *Cell* 128 (3), 445-458 (2007).

[9] Radtke, F. & Clevers, H., Self-renewal and cancer of the gut: two sides of a coin. *Science* 307 (5717), 1904-1909 (2005).

[10] McDonald, S.A., Preston, S.L., Lovell, M.J., Wright, N.A., & Jankowski, J.A., Mechanisms of disease: from stem cells to colorectal cancer. *Nat Clin Pract Gastroenterol Hepatol* 3 (5), 267-274 (2006).

[11] Spradling, A., Drummond-Barbosa, D., & Kai, T., Stem cells find their niche. *Nature* 414 (6859), 98-104 (2001).

[12] Sangiorgi, E. & Capecchi, M.R., Bmi1 is expressed in vivo in intestinal stem cells. *Nat Genet* 40 (7), 915-920 (2008).

[13] Potten, C.S., Kovacs, L., & Hamilton, E., Continuous labelling studies on mouse skin and intestine. *Cell Tissue Kinet* 7 (3), 271-283 (1974).

[14] Bjerknes, M. & Cheng, H., The stem-cell zone of the small intestinal epithelium. V. Evidence for controls over orientation of boundaries between the stem-cell zone, proliferative zone, and the maturation zone. *Am J Anat* 160 (1), 105-112 (1981).

[15] Cheng, H. & Leblond, C.P., Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. III. Entero-endocrine cells. *Am J Anat* 141 (4), 503-519 (1974).

[16] Leedham, S.J., Brittan, M., McDonald, S.A., & Wright, N.A., Intestinal stem cells. *J Cell Mol Med* 9 (1), 11-24 (2005).

[17] Batlle E, Henderson JT, Beghtel H, van den Born MM, Sancho E, Huls G, Meeldijk J, Robertson J, van de Wetering M, Pawson T, Clevers H. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* 111 (2), 251-263 (2002).

[18] Potten, C.S., Kellett, M., Roberts, S.A., Rew, D.A., & Wilson, G.D., Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* 33 (1), 71-78 (1992).

[19] Booth, C. & Potten, C.S., Gut instincts: thoughts on intestinal epithelial stem cells. *J Clin Invest* 105 (11), 1493-1499 (2000).

[20] Nishimura, S., Wakabayashi, N., Toyoda, K., Kashima, K., & Mitsufuji, S., Expression of Musashi-1 in human normal colon crypt cells: a possible stem cell marker of human colon epithelium. *Dig Dis Sci* 48 (8), 1523-1529 (2003).

[21] Imai T, Tokunaga A, Yoshida T, Hashimoto M, Mikoshiba K, Weinmaster G, Nakafuku M, Okano H. The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol* 21 (12), 3888-3900 (2001).

[22] Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, Clevers H. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449 (7165), 1003-1007 (2007).

[23] Zhu L, Gibson P, Currle DS, Tong Y, Richardson RJ, Bayazitov IT, Poppleton H, Zakharenko S, Ellison DW, Gilbertson RJ. Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* 457 (7229), 603-607 (2009).

[24] Fujimoto, K., Beauchamp, R.D., & Whitehead, R.H., Identification and isolation of candidate human colonic clonogenic cells based on cell surface integrin expression. *Gastroenterology* 123 (6), 1941-1948 (2002).

[25] Holmberg J, Genander M, Halford MM, Annerén C, Sondell M, Chumley MJ, Silvany RE, Henkemeyer M, Frisén J. EphB receptors coordinate migration and proliferation in the intestinal stem cell niche. *Cell* 125 (6), 1151-1163 (2006).

[26] Heppner, G.H., Tumor heterogeneity. *Cancer Res* 44 (6), 2259-2265 (1984).

[27] Hamburger, A. & Salmon, S.E., Primary bioassay of human myeloma stem cells. *J Clin Invest* 60 (4), 846-854 (1977).

[28] Jamieson CH, Ailles LE, Dylla SJ, Muijtjens M, Jones C, Zehnder JL, Gotlib J, Li K, Manz MG, Keating A, Sawyers CL, Weissman IL. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med* 351 (7), 657-667 (2004).

[29] Shih IM, Wang TL, Traverso G, Romans K, Hamilton SR, Ben-Sasson S, Kinzler KW, Vogelstein B. Top-down morphogenesis of colorectal tumors. *Proc Natl Acad Sci U S A* 98 (5), 2640-2645 (2001).

[30] He, S., Nakada, D., & Morrison, S.J., Mechanisms of stem cell self-renewal. *Annu Rev Cell Dev Biol* 25, 377-406 (2009).

[31] Kelly, P.N., Dakic, A., Adams, J.M., Nutt, S.L., & Strasser, A., Tumor growth need not be driven by rare cancer stem cells. *Science* 317 (5836), 337 (2007).

[32] Fodde, R. & Brabletz, T., Wnt/beta-catenin signaling in cancer stemness and malignant behavior. *Curr Opin Cell Biol* 19 (2), 150-158 (2007).

[33] Fearon, E.R. & Vogelstein, B., A genetic model for colorectal tumorigenesis. *Cell* 61 (5), 759-767 (1990).

[34] Batlle E, Bacani J, Begthel H, Jonkheer S, Gregorieff A, van de Born M, Malats N, Sancho E, Boon E, Pawson T, Gallinger S, Pals S, Clevers H. EphB receptor activity suppresses colorectal cancer progression. *Nature* 435 (7045), 1126-1130 (2005).

[35] Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, Danenberg E, Clarke AR, Sansom OJ, Clevers H. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* 457 (7229), 608-611 (2009).

[36] Brabletz T, Jung A, Reu S, Porzner M, Hlubek F, Kunz-Schughart LA, Knuechel R, Kirchner T. Variable beta-catenin expression in colorectal cancers indicates

tumor progression driven by the tumor environment. *Proc Natl Acad Sci U S A* 98 (18), 10356-10361 (2001).

[37] Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De Maria R. Identification and expansion of human colon-cancer-initiating cells. *Nature* 445 (7123), 111-115 (2007).

[38] O'Brien, C.A., Pollett, A., Gallinger, S., & Dick, J.E., A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445 (7123), 106-110 (2007).

[39] Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, Hoey T, Gurney A, Huang EH, Simeone DM, Shelton AA, Parmiani G, Castelli C, Clarke MF. Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A* 104 (24), 10158-10163 (2007).

[40] Shmelkov SV, Butler JM, Hooper AT, Hormigo A, Kushner J, Milde T, St Clair R, Baljevic M, White I, Jin DK, Chadburn A, Murphy AJ, Valenzuela DM, Gale NW, Thurston G, Yancopoulos GD, D'Angelica M, Kemeny N, Lyden D, Rafii S. CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. *J Clin Invest* 118 (6), 2111-2120 (2008).

[41] Snippert HJ, van Es JH, van den Born M, Begthel H, Stange DE, Barker N, Clevers H. Prominin-1/CD133 marks stem cells and early progenitors in mouse small intestine. *Gastroenterology* 136 (7), 2187-2194 e2181 (2009).

[42] Vermeulen L, De Sousa E Melo F, van der Heijden M, Cameron K, de Jong JH, Borovski T, Tuynman JB, Todaro M, Merz C, Rodermond H, Sprick MR, Kemper K, Richel DJ, Stassi G, Medema JP. Wnt activity defines colon cancer

stem cells and is regulated by the microenvironment. *Nat Cell Biol* 12 (5), 468-476.

[43] Willert K, Brown JD, Danenberg E, Duncan AW, Weissman IL, Reya T, Yates JR 3rd, Nusse R. Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* 423 (6938), 448-452 (2003).

[44] Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, Vogelstein B, Kinzler KW. APC mutations occur early during colorectal tumorigenesis. *Nature* 359 (6392), 235-237 (1992).

[45] Miyoshi Y, Ando H, Nagase H, Nishisho I, Horii A, Miki Y, Mori T, Utsunomiya J, Baba S, Petersen G, *et al.* Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proc Natl Acad Sci U S A* 89 (10), 4452-4456 (1992).

[46] Phelps, R.A., Broadbent, T.J., Stafforini, D.M., & Jones, D.A., New perspectives on APC control of cell fate and proliferation in colorectal cancer. *Cell Cycle* 8 (16), 2549-2556 (2009).

[47] Bienz, M. & Clevers, H., Linking colorectal cancer to Wnt signaling. *Cell* 103 (2), 311-320 (2000).

[48] Clevers, H., Wnt/beta-catenin signaling in development and disease. *Cell* 127 (3), 469-480 (2006).

[49] Spink, K.E., Polakis, P., & Weis, W.I., Structural basis of the Axin-adenomatous polyposis coli interaction. *EMBO J* 19 (10), 2270-2279 (2000).

[50] Polakis, P., Casein kinase 1: a Wnt'er of disconnect. *Curr Biol* 12 (14), R499-R501 (2002).

[51] Rubinfeld, B., Tice, D.A., & Polakis, P., Axin-dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1epsilon. *J Biol Chem* 276 (42), 39037-39045 (2001).

[52] Neufeld, K.L., Zhang, F., Cullen, B.R., & White, R.L., APC-mediated downregulation of beta-catenin activity involves nuclear sequestration and nuclear export. *EMBO Rep* 1 (6), 519-523 (2000).

[53] Rosin-Arbesfeld, R., Townsley, F., & Bienz, M., The APC tumour suppressor has a nuclear export function. *Nature* 406 (6799), 1009-1012 (2000).

[54] Wong, N.A. & Pignatelli, M., Beta-catenin--a linchpin in colorectal carcinogenesis? *Am J Pathol* 160 (2), 389-401 (2002).

[55] Moon, R.T., Bowerman, B., Boutros, M., & Perrimon, N., The promise and perils of Wnt signaling through beta-catenin. *Science* 296 (5573), 1644-1646 (2002).

[56] Murray NR, Davidson LA, Chapkin RS, Clay Gustafson W, Schattenberg DG, Fields AP. Overexpression of protein kinase C betaII induces colonic hyperproliferation and increased sensitivity to colon carcinogenesis. *J Cell Biol* 145 (4), 699-711 (1999).

[57] van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111 (2), 241-250 (2002).

[58] Mann B, Gelos M, Siedow A, Hanski ML, Gratchev A, Ilyas M, Bodmer WF, Moyer MP, Riecken EO, Buhr HJ, Hanski C. Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. *Proc Natl Acad Sci U S A* 96 (4), 1603-1608 (1999).

[59] Kajiguchi T, Chung EJ, Lee S, Stine A, Kiyoi H, Naoe T, Levis MJ, Neckers L, Trepel JB. FLT3 regulates beta-catenin tyrosine phosphorylation, nuclear localization, and transcriptional activity in acute myeloid leukemia cells. *Leukemia* 21 (12), 2476-2484 (2007).

[60] Kajiguchi, T., Lee, S., Lee, M.J., Trepel, J.B., & Neckers, L., KIT regulates tyrosine phosphorylation and nuclear localization of beta-catenin in mast cell leukemia. *Leuk Res* 32 (5), 761-770 (2008).

[61] Shtutman, M., Zhurinsky, J., Oren, M., Levina, E., & Ben-Ze'ev, A., PML is a target gene of beta-catenin and plakoglobin, and coactivates beta-catenin-mediated transcription. *Cancer Res* 62 (20), 5947-5954 (2002).

[62] Roura, S., Miravet, S., Piedra, J., Garcia de Herreros, A., & Dunach, M., Regulation of E-cadherin/Catenin association by tyrosine phosphorylation. *J Biol Chem* 274 (51), 36734-36740 (1999).

[63] Harris, T.J. & Tepass, U., Adherens junctions: from molecules to morphogenesis. *Nat Rev Mol Cell Biol* 11 (7), 502-514.

[64] Deng J, Miller SA, Wang HY, Xia W, Wen Y, Zhou BP, Li Y, Lin SY, Hung MC. beta-catenin interacts with and inhibits NF-kappa B in human colon and breast cancer. *Cancer Cell* 2 (4), 323-334 (2002).

[65] Steigerwald, K., Behbehani, G.K., Combs, K.A., Barton, M.C., & Groden, J., The APC tumor suppressor promotes transcription-independent apoptosis in vitro. *Mol Cancer Res* 3 (2), 78-89 (2005).

[66] Pinto, D., Gregorieff, A., Begthel, H., & Clevers, H., Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev* 17 (14), 1709-1713 (2003).

[67] van Es JH, van Gijn ME, Riccio O, van den Born M, Vooijs M, Begthel H, Cozijnsen M, Robine S, Winton DJ, Radtke F, Clevers H. Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* 435 (7044), 959-963 (2005).

[68] Bos, J.L., ras oncogenes in human cancer: a review. *Cancer Res* 49 (17), 4682-4689 (1989).

[69] Andreyev, H.J., Norman, A.R., Cunningham, D., Oates, J.R., & Clarke, P.A., Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst* 90 (9), 675-684 (1998).

[70] Kerkhoff, E. & Rapp, U.R., Cell cycle targets of Ras/Raf signalling. *Oncogene* 17 (11 Reviews), 1457-1462 (1998).

[71] Fang, J.Y. & Richardson, B.C., The MAPK signalling pathways and colorectal cancer. *Lancet Oncol* 6 (5), 322-327 (2005).

[72] Fujita T, Washio K, Takabatake D, Takahashi H, Yoshitomi S, Tsukuda K, Ishibe Y, Ogasawara Y, Doihara H, Shimizu N. Proteasome inhibitors can alter the signaling pathways and attenuate the P-glycoprotein-mediated multidrug resistance. *Int J Cancer* 117 (4), 670-682 (2005).

[73] Schepers, H., Wierenga, A.T., Eggen, B.J., & Vellenga, E., Oncogenic Ras blocks transforming growth factor-beta-induced cell-cycle arrest by degradation of p27 through a MEK/Erk/SKP2-dependent pathway. *Exp Hematol* 33 (7), 747-757 (2005).

[74] Bhattacharya, S., Mathew, G., Jayne, D.G., Pelengaris, S., & Khan, M., 15-Lipoxygenase-1 in Colorectal Cancer: A Review. *Tumour Biol* 30 (4), 185-199 (2009).

[75] Cardone MH, Roy N, Stennicke HR, Salvesen GS, Franke TF, Stanbridge E, Frisch S, Reed JC. Regulation of cell death protease caspase-9 by phosphorylation. *Science* 282 (5392), 1318-1321 (1998).

[76] Aoki, M., Jiang, H., & Vogt, P.K., Proteasomal degradation of the FoxO1 transcriptional regulator in cells transformed by the P3k and Akt oncoproteins. *Proc Natl Acad Sci U S A* 101 (37), 13613-13617 (2004).

[77] Thomas, G.V., mTOR and cancer: reason for dancing at the crossroads? *Curr Opin Genet Dev* 16 (1), 78-84 (2006).

[78] Vivanco, I. & Sawyers, C.L., The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev Cancer* 2 (7), 489-501 (2002).

[79] Cantley, L.C. & Neel, B.G., New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc Natl Acad Sci U S A* 96 (8), 4240-4245 (1999).

[80] Cullen, P.J. & Lockyer, P.J., Integration of calcium and Ras signalling. *Nat Rev Mol Cell Biol* 3 (5), 339-348 (2002).

[81] Hayward, P., Kalmar, T., & Arias, A.M., Wnt/Notch signalling and information processing during development. *Development* 135 (3), 411-424 (2008).

[82] Haines, N. & Irvine, K.D., Glycosylation regulates Notch signalling. *Nat Rev Mol Cell Biol* 4 (10), 786-797 (2003).

[83] Cohen B, Bashirullah A, Dagnino L, Campbell C, Fisher WW, Leow CC, Whiting E, Ryan D, Zinyk D, Boulianne G, Hui CC, Gallie B, Phillips RA, Lipshitz HD, Egan SE. Fringe boundaries coincide with Notch-dependent patterning centres in mammals and alter Notch-dependent development in Drosophila. *Nat Genet* 16 (3), 283-288 (1997).

[84] Johnston SH, Rauskolb C, Wilson R, Prabhakaran B, Irvine KD, Vogt TF. A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway. *Development* 124 (11), 2245-2254 (1997).

[85] Kopan, R., Notch: a membrane-bound transcription factor. *J Cell Sci* 115 (Pt 6), 1095-1097 (2002).

[86] Yang, Q., Bermingham, N.A., Finegold, M.J., & Zoghbi, H.Y., Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* 294 (5549), 2155-2158 (2001).

[87] Riccio O, van Gijn ME, Bezdek AC, Pellegrinet L, van Es JH, Zimber-Strobl U, Strobl LJ, Honjo T, Clevers H, Radtke F. Loss of intestinal crypt progenitor cells owing to inactivation of both Notch1 and Notch2 is accompanied by derepression of CDK inhibitors p27Kip1 and p57Kip2. *EMBO Rep* 9 (4), 377-383 (2008).

[88] Reedijk M, Odorcic S, Zhang H, Chetty R, Tennert C, Dickson BC, Lockwood G, Gallinger S, Egan SE. Activation of Notch signaling in human colon adenocarcinoma. *Int J Oncol* 33 (6), 1223-1229 (2008).

[89] Leow, C.C., Romero, M.S., Ross, S., Polakis, P., & Gao, W.Q., Hath1, down-regulated in colon adenocarcinomas, inhibits proliferation and tumorigenesis of colon cancer cells. *Cancer Res* 64 (17), 6050-6057 (2004).

[90] Leow, C.C., Polakis, P., & Gao, W.Q., A role for Hath1, a bHLH transcription factor, in colon adenocarcinoma. *Ann N Y Acad Sci* 1059, 174-183 (2005).

[91] Herpin, A., Lelong, C., & Favrel, P., Transforming growth factor-beta-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. *Dev Comp Immunol* 28 (5), 461-485 (2004).

[92] Khalil, N., TGF-beta: from latent to active. *Microbes Infect* 1 (15), 1255-1263 (1999).

[93] Daopin, S., Piez, K.A., Ogawa, Y., & Davies, D.R., Crystal structure of transforming growth factor-beta 2: an unusual fold for the superfamily. *Science* 257 (5068), 369-373 (1992).

[94] Attisano, L. & Wrana, J.L., Signal transduction by the TGF-beta superfamily. *Science* 296 (5573), 1646-1647 (2002).

[95] Hanahan, D. & Weinberg, R.A., The hallmarks of cancer. *Cell* 100 (1), 57-70 (2000).

[96] Tadlock, L., Yamagiwa, Y., Hawker, J., Marienfeld, C., & Patel, T., Transforming growth factor-beta inhibition of proteasomal activity: a potential mechanism of growth arrest. *Am J Physiol Cell Physiol* 285 (2), C277-285 (2003).

[97] Yan, X., Liu, Z., & Chen, Y., Regulation of TGF-beta signaling by Smad7. *Acta Biochim Biophys Sin (Shanghai)* 41 (4), 263-272 (2009).

[98] Wu, M.Y. & Hill, C.S., Tgf-beta superfamily signaling in embryonic development and homeostasis. *Dev Cell* 16 (3), 329-343 (2009).

[99] Eppert K, Scherer SW, Ozcelik H, Pirone R, Hoodless P, Kim H, Tsui LC, Bapat B, Gallinger S, Andrulis IL, Thomsen GH, Wrana JL, Attisano L. MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* 86 (4), 543-552 (1996).

[100] Miyaki M, Iijima T, Konishi M, Sakai K, Ishii A, Yasuno M, Hishima T, Koike M, Shitara N, Iwama T, Utsunomiya J, Kuroki T, Mori T. Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* 18 (20), 3098-3103 (1999).

[101] Wakefield, L.M. & Roberts, A.B., TGF-beta signaling: positive and negative effects on tumorigenesis. *Curr Opin Genet Dev* 12 (1), 22-29 (2002).

[102] Elliott, R.L. & Blobe, G.C., Role of transforming growth factor Beta in human cancer. *J Clin Oncol* 23 (9), 2078-2093 (2005).

[103] Saha, D., Datta, P.K., & Beauchamp, R.D., Oncogenic ras represses transforming growth factor-beta /Smad signaling by degrading tumor suppressor Smad4. *J Biol Chem* 276 (31), 29531-29537 (2001).

[104] Lo, R.S., Wotton, D., & Massague, J., Epidermal growth factor signaling via Ras controls the Smad transcriptional co-repressor TGIF. *EMBO J* 20 (1-2), 128-136 (2001).

[105] Furuhashi M, Yagi K, Yamamoto H, Furukawa Y, Shimada S, Nakamura Y, Kikuchi A, Miyazono K, Kato M. Axin facilitates Smad3 activation in the transforming growth factor beta signaling pathway. *Mol Cell Biol* 21 (15), 5132-5141 (2001).

[106] Labbe, E., Letamendia, A., & Attisano, L., Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. *Proc Natl Acad Sci U S A* 97 (15), 8358-8363 (2000).

[107] Chen, D., Zhao, M., & Mundy, G.R., Bone Morphogenetic Proteins. *Growth Factors* 22 (4), 233-241 (2004).

[108] Urist, M.R., Bone: formation by autoinduction. *Science* 150 (698), 893-899 (1965).

[109] Wozney, J.M., The bone morphogenetic protein family and osteogenesis. *Mol Reprod Dev* 32 (2), 160-167 (1992).

[110] Wozney JM, Rosen V, Celeste AJ, Mitsock LM, Whitters MJ, Kriz RW, Hewick RM, Wang EA. Novel regulators of bone formation: molecular clones and activities. *Science* 242 (4885), 1528-1534 (1988).

[111] Luyten FP, Cunningham NS, Ma S, Muthukumaran N, Hammonds RG, Nevins WB, Woods WI, Reddi AH. Purification and partial amino acid sequence of osteogenin, a protein initiating bone differentiation. *J Biol Chem* 264 (23), 13377-13380 (1989).

[112] Munoz-Sanjuan, I. & Brivanlou, A.H., Neural induction, the default model and embryonic stem cells. *Nat Rev Neurosci* 3 (4), 271-280 (2002).

[113] Christiansen, J.H., Coles, E.G., & Wilkinson, D.G., Molecular control of neural crest formation, migration and differentiation. *Curr Opin Cell Biol* 12 (6), 719-724 (2000).

[114] Wall, N.A. & Hogan, B.L., TGF-beta related genes in development. *Curr Opin Genet Dev* 4 (4), 517-522 (1994).

[115] Koenig BB, Cook JS, Wolsing DH, Ting J, Tiesman JP, Correa PE, Olson CA, Pecquet AL, Ventura F, Grant RA, *et al*, Characterization and cloning of a receptor for BMP-2 and BMP-4 from NIH 3T3 cells. *Mol Cell Biol* 14 (9), 5961-5974 (1994).

[116] ten Dijke P, Yamashita H, Sampath TK, Reddi AH, Estevez M, Riddle DL, Ichijo H, Heldin CH, Miyazono K. Identification of type I receptors for osteogenic protein-1 and bone morphogenetic protein-4. *J Biol Chem* 269 (25), 16985-16988 (1994).

[117] Macias-Silva, M., Hoodless, P.A., Tang, S.J., Buchwald, M., & Wrana, J.L., Specific activation of Smad1 signaling pathways by the BMP7 type I receptor, ALK2. *J Biol Chem* 273 (40), 25628-25636 (1998).

[118] Yamashita H, ten Dijke P, Huylebroeck D, Sampath TK, Andries M, Smith JC, Heldin CH, Miyazono K.Osteogenic protein-1 binds to activin type II receptors and induces certain activin-like effects. *J Cell Biol* 130 (1), 217-226 (1995).

[119] Rosenzweig BL, Imamura T, Okadome T, Cox GN, Yamashita H, ten Dijke P, Heldin CH, Miyazono K. Cloning and characterization of a human type II receptor for bone morphogenetic proteins. *Proc Natl Acad Sci U S A* 92 (17), 7632-7636 (1995).

[120] Kawabata, M., Chytil, A., & Moses, H.L., Cloning of a novel type II serine/threonine kinase receptor through interaction with the type I transforming growth factor-beta receptor. *J Biol Chem* 270 (10), 5625-5630 (1995).

[121] Moustakas, A. & Heldin, C.H., From mono- to oligo-Smads: the heart of the matter in TGF-beta signal transduction. *Genes Dev* 16 (15), 1867-1871 (2002).

[122] Hoodless PA, Haerry T, Abdollah S, Stapleton M, O'Connor MB, Attisano L, Wrana JL. MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* 85 (4), 489-500 (1996).

[123] Chen, Y., Bhushan, A., & Vale, W., Smad8 mediates the signaling of the ALK-2 [corrected] receptor serine kinase. *Proc Natl Acad Sci U S A* 94 (24), 12938-12943 (1997).

[124] Nishimura R, Kato Y, Chen D, Harris SE, Mundy GR, Yoneda T.Smad5 and DPC4 are key molecules in mediating BMP-2-induced osteoblastic differentiation of the pluripotent mesenchymal precursor cell line C2C12. *J Biol Chem* 273 (4), 1872-1879 (1998).

[125] Devlin RD, Du Z, Pereira RC, Kimble RB, Economides AN, Jorgetti V, Canalis E. Skeletal overexpression of noggin results in osteopenia and reduced bone formation. *Endocrinology* 144 (5), 1972-1978 (2003).

[126] Imamura T, Takase M, Nishihara A, Oeda E, Hanai J, Kawabata M, Miyazono K. Smad6 inhibits signalling by the TGF-beta superfamily. *Nature* 389 (6651), 622-626 (1997).

[127] Yoshida Y, Tanaka S, Umemori H, Minowa O, Usui M, Ikematsu N, Hosoda E, Imamura T, Kuno J, Yamashita T, Miyazono K, Noda M, Noda T, Yamamoto T. Negative regulation of BMP/Smad signaling by Tob in osteoblasts. *Cell* 103 (7), 1085-1097 (2000).

[128] Dicuonzo G, Angeletti S, Garcia-Foncillas J, Brugarolas A, Okrouzhnov Y, Santini D, Tonini G, Lorino G, De Cesaris M, Baldi A. Colorectal carcinomas and PTEN/MMAC1 gene mutations. *Clin Cancer Res* 7 (12), 4049-4053 (2001).

[129] Kleeff J, Maruyama H, Ishiwata T, Sawhney H, Friess H, Büchler MW, Korc M. Bone morphogenetic protein 2 exerts diverse effects on cell growth in vitro and is expressed in human pancreatic cancer in vivo. *Gastroenterology* 116 (5), 1202-1216 (1999).

[130] Moskaluk CA, Hruban RH, Schutte M, Lietman AS, Smyrk T, Fusaro L, Fusaro R, Lynch J, Yeo CJ, Jackson CE, Lynch HT, Kern SE. Genomic sequencing of DPC4 in the analysis of familial pancreatic carcinoma. *Diagn Mol Pathol* 6 (2), 85-90 (1997).

[131] Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. Gene expression profiles in normal and cancer cells. *Science* 276 (5316), 1268-1272 (1997).

[132] Hardwick JC, Van Den Brink GR, Bleuming SA, Ballester I, Van Den Brande JM, Keller JJ, Offerhaus GJ, Van Deventer SJ, Peppelenbosch MP. Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon. *Gastroenterology* 126 (1), 111-121 (2004).

[133] Howe JR, Roth S, Ringold JC, Summers RW, Järvinen HJ, Sistonen P, Tomlinson IP, Houlston RS, Bevan S, Mitros FA, Stone EM, Aaltonen LA.

Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science* 280 (5366), 1086-1088 (1998).

134 Zhou XP, Woodford-Richens K, Lehtonen R, Kurose K, Aldred M, Hampel H, Launonen V, Virta S, Pilarski R, Salovaara R, Bodmer WF, Conrad BA, Dunlop M, Hodgson SV, Iwama T, Järvinen H, Kellokumpu I, Kim JC, Leggett B, Markie D, Mecklin JP, Neale K, Phillips R, Piris J, Rozen P, Houlston RS, Aaltonen LA, Tomlinson IP, Eng C. Germline mutations in BMPR1A/ALK3 cause a subset of cases of juvenile polyposis syndrome and of Cowden and Bannayan-Riley-Ruvalcaba syndromes. *Am J Hum Genet* 69 (4), 704-711 (2001).

135 Howe JR, Sayed MG, Ahmed AF, Ringold J, Larsen-Haidle J, Merg A, Mitros FA, Vaccaro CA, Petersen GM, Giardiello FM, Tinley ST, Aaltonen LA, Lynch HT. The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations. *J Med Genet* 41 (7), 484-491 (2004).

136 Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW, *et al*. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* 247 (4938), 49-56 (1990).

137 Loh K, Chia JA, Greco S, Cozzi SJ, Buttenshaw RL, Bond CE, Simms LA, Pike T, Young JP, Jass JR, Spring KJ, Leggett BA, Whitehall VL. Bone morphogenetic protein signaling and growth suppression in colon cancer. *Am J Physiol Gastrointest Liver Physiol* 291 (1), G135-145 (2006).

138 He XC, Zhang J, Tong WG, Tawfik O, Ross J, Scoville DH, Tian Q, Zeng X, He X, Wiedemann LM, Mishina Y, Li L. BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat Genet* 36 (10), 1117-1121 (2004).

[139] Haramis AP, Begthel H, van den Born M, van Es J, Jonkheer S, Offerhaus GJ, Clevers H. De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* 303 (5664), 1684-1686 (2004).

[140] Sancho, E., Batlle, E., & Clevers, H., Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol* 20, 695-723 (2004).

[141] Lombardo Y, Scopelliti A, Cammareri P, Todaro M, Iovino F, Ricci-Vitiani L, Gulotta G, Dieli F, de Maria R, Stassi G. Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. *Gastroenterology* 140 (1), 297-309.

[142] Coutts, A.S. & La Thangue, N.B., The p53 response: emerging levels of co-factor complexity. *Biochem Biophys Res Commun* 331 (3), 778-785 (2005).

[143] Liu J, Stevens J, Rote CA, Yost HJ, Hu Y, Neufeld KL, White RL, Matsunami N. Siah-1 mediates a novel beta-catenin degradation pathway linking p53 to the adenomatous polyposis coli protein. *Mol Cell* 7 (5), 927-936 (2001).

[144] Iacopetta, B., TP53 mutation in colorectal cancer. *Hum Mutat* 21 (3), 271-276 (2003).

[145] Kohn, K.W. & Pommier, Y., Molecular interaction map of the p53 and Mdm2 logic elements, which control the Off-On switch of p53 in response to DNA damage. *Biochem Biophys Res Commun* 331 (3), 816-827 (2005).

[146] Dornan D, Wertz I, Shimizu H, Arnott D, Frantz GD, Dowd P, O'Rourke K, Koeppen H, Dixit VM. The ubiquitin ligase COP1 is a critical negative regulator of p53. *Nature* 429 (6987), 86-92 (2004).

[147] Leng RP, Lin Y, Ma W, Wu H, Lemmers B, Chung S, Parant JM, Lozano G, Hakem R, Benchimol S. Pirh2, a p53-induced ubiquitin-protein ligase, promotes p53 degradation. *Cell* 112 (6), 779-791 (2003).

[148] Chen D, Kon N, Li M, Zhang W, Qin J, Gu W. ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor. *Cell* 121 (7), 1071-1083 (2005).

[149] Gong Z, Hebert JR, Bostick RM, Deng Z, Hurley TG, Dixon DA, Nitcheva D, Xie D. Common polymorphisms in 5-lipoxygenase and 12-lipoxygenase genes and the risk of incident, sporadic colorectal adenoma. *Cancer* 109 (5), 849-857 (2007).

[150] Shureiqi, I. & Lippman, S.M., Lipoxygenase modulation to reverse carcinogenesis. *Cancer Res* 61 (17), 6307-6312 (2001).

[151] Hamberg, M. & Samuelsson, B., Prostaglandin endoperoxides. Novel transformations of arachidonic acid in human platelets. *Proc Natl Acad Sci U S A* 71 (9), 3400-3404 (1974).

[152] Schewe, T., Halangk, W., Hiebsch, C., & Rapoport, S.M., A lipoxygenase in rabbit reticulocytes which attacks phospholipids and intact mitochondria. *FEBS Lett* 60 (1), 149-152 (1975).

[153] Brash, A.R., Lipoxygenases: occurrence, functions, catalysis, and acquisition of substrate. *J Biol Chem* 274 (34), 23679-23682 (1999).

[154] Nixon, J.B., Kim, K.S., Lamb, P.W., Bottone, F.G., & Eling, T.E., 15-Lipoxygenase-1 has anti-tumorigenic effects in colorectal cancer. *Prostaglandins Leukot Essent Fatty Acids* 70 (1), 7-15 (2004).

[155] Shureiqi I, Jiang W, Zuo X, Wu Y, Stimmel JB, Leesnitzer LM, Morris JS, Fan HZ, Fischer SM, Lippman SM. The 15-lipoxygenase-1 product 13-S-hydroxyoctadecadienoic acid down-regulates PPAR-delta to induce apoptosis in colorectal cancer cells. *Proc Natl Acad Sci U S A* 100 (17), 9968-9973 (2003).

[156] Wu Y, Fang B, Yang XQ, Wang L, Chen D, Krasnykh V, Carter BZ, Morris JS, Shureiqi I. Therapeutic molecular targeting of 15-lipoxygenase-1 in colon cancer. *Mol Ther* 16 (5), 886-892 (2008).

[157] Yoshinaga, M., Buchanan, F.G., & DuBois, R.N., 15-LOX-1 inhibits p21 (Cip/WAF 1) expression by enhancing MEK-ERK 1/2 signaling in colon carcinoma cells. *Prostaglandins Other Lipid Mediat* 73 (1-2), 111-122 (2004).

[158] Hsi, L.C., Wilson, L., Nixon, J., & Eling, T.E., 15-lipoxygenase-1 metabolites down-regulate peroxisome proliferator-activated receptor gamma via the MAPK signaling pathway. *J Biol Chem* 276 (37), 34545-34552 (2001).

[159] Faca, V., Krasnoselsky, A., & Hanash, S., Innovative proteomic approaches for cancer biomarker discovery. *Biotechniques* 43 (3), 279, 281-273, 285 (2007).

[160] Kumar, S., Mohan, A., & Guleria, R., Biomarkers in cancer screening, research and detection: present and future: a review. *Biomarkers* 11 (5), 385-405 (2006).

[161] Pavlou, M.P. & Diamandis, E.P., The cancer cell secretome: a good source for discovering biomarkers? *J Proteomics* 73 (10), 1896-1906.

[162] Hanash, S.M., Pitteri, S.J., & Faca, V.M., Mining the plasma proteome for cancer biomarkers. *Nature* 452 (7187), 571-579 (2008).

[163] Ludwig, J.A. & Weinstein, J.N., Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 5 (11), 845-856 (2005).

[164] Petricoin, E., Wulfkuhle, J., Espina, V., & Liotta, L.A., Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. *J Proteome Res* 3 (2), 209-217 (2004).

[165] Cho, W.C., Contribution of oncoproteomics to cancer biomarker discovery. *Mol Cancer* 6, 25 (2007).

[166] Alaiya, A., Al-Mohanna, M., & Linder, S., Clinical cancer proteomics: promises and pitfalls. *J Proteome Res* 4 (4), 1213-1222 (2005).

[167] Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 93 (14), 1054-1061 (2001).

[168] Hudler, P., Gorsic, M., & Komel, R., Proteomic strategies and challenges in tumor metastasis research. *Clin Exp Metastasis* 27 (6), 441-451.

[169] Zieglschmid, V., Hollmann, C., & Bocher, O., Detection of disseminated tumor cells in peripheral blood. *Crit Rev Clin Lab Sci* 42 (2), 155-196 (2005).

[170] Omenn, G.S., Strategies for plasma proteomic profiling of cancers. *Proteomics* 6 (20), 5662-5673 (2006).

[171] Paterlini-Brechot, P. & Benali, N.L., Circulating tumor cells (CTC) detection: clinical impact and future directions. *Cancer Lett* 253 (2), 180-204 (2007).

[172] Lu M, Faull KF, Whitelegge JP, He J, Shen D, Saxton RE, Chang HR. Proteomics and mass spectrometry for cancer biomarker discovery. *Biomark Insights* 2, 347-360 (2007).

[173] Mischak H, Coon JJ, Novak J, Weissinger EM, Schanstra JP, Dominiczak AF. Capillary electrophoresis-mass spectrometry as a powerful tool in biomarker discovery and clinical diagnosis: an update of recent developments. *Mass Spectrom Rev* 28 (5), 703-724 (2009).

[174] Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X, Wang R, Wasinger V, Wu CY, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A, Pisano M, Andrews

P, Tammen H, Speicher DW, Hanash SM. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 5 (13), 3226-3245 (2005).

[175] de la Grange, P., Gratadou, L., Delord, M., Dutertre, M., & Auboeuf, D., Splicing factor and exon profiling across human tissues. *Nucleic Acids Res* 38 (9), 2825-2838.

[176] Dutertre M, Lacroix-Triki M, Driouch K, de la Grange P, Gratadou L, Beck S, Millevoi S, Tazi J, Lidereau R, Vagner S, Auboeuf D. Exon-based clustering of murine breast tumor transcriptomes reveals alternative exons whose expression is associated with metastasis. *Cancer Res* 70 (3), 896-905.

[177] Qu, X. & Komatsu, T., Molecular capture in protein nanotubes. *ACS Nano* 4 (1), 563-573.

[178] Warder SE, Tucker LA, Strelitzer TJ, McKeegan EM, Meuth JL, Jung PM, Saraf A, Singh B, Lai-Zhang J, Gagne G, Rogers JC. Reducing agent-mediated precipitation of high-abundance plasma proteins. *Anal Biochem* 387 (2), 184-193 (2009).

[179] Fang, X. & Zhang, W.W., Affinity separation and enrichment methods in proteomic analysis. *J Proteomics* 71 (3), 284-303 (2008).

[180] Polaskova, V., Kapur, A., Khan, A., Molloy, M.P., & Baker, M.S., High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. *Electrophoresis* 31 (3), 471-482.

[181] Liotta, L.A. & Petricoin, E.F., Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest* 116 (1), 26-30 (2006).

[182] Lowenthal MS, Mehta AI, Frogale K, Bandle RW, Araujo RP, Hood BL, Veenstra TD, Conrads TP, Goldsmith P, Fishman D, Petricoin EF 3rd, Liotta LA. Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin Chem* 51 (10), 1933-1945 (2005).

[183] Lai, X., Liangpunsakul, S., Crabb, D.W., Ringham, H.N., & Witzmann, F.A., A proteomic workflow for discovery of serum carrier protein-bound biomarker candidates of alcohol abuse using LC-MS/MS. *Electrophoresis* 30 (12), 2207-2214 (2009).

[184] Lopez MF, Mikulskis A, Kuzdzal S, Bennett DA, Kelly J, Golenko E, DiCesare J, Denoyer E, Patton WF, Ediger R, Sapp L, Ziegert T, Lynch C, Kramer S, Whiteley GR, Wall MR, Mannion DP, Della Cioppa G, Rakitan JS, Wolfe GM. High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific, carrier-protein-bound mass signatures. *Clin Chem* 51 (10), 1946-1954 (2005).

[185] Hortin, G.L., The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin Chem* 52 (7), 1223-1237 (2006).

[186] Fredolini C, Tamburro D, Gambara G, Lepene BS, Espina V, Petricoin EF 3rd, Liotta LA, Luchini A. Nanoparticle technology: amplifying the effective sensitivity of biomarker detection to create a urine test for hGH. *Drug Test Anal* 1 (9-10), 447-454 (2009).

[187] Luchini A, Fredolini C, Espina BH, Meani F, Reeder A, Rucker S, Petricoin EF 3rd, Liotta LA. Nanoparticle technology: addressing the fundamental roadblocks to protein biomarker discovery. *Curr Mol Med* 10 (2), 133-141.

[188] Kolch, W., Mischak, H., & Pitt, A.R., The molecular make-up of a tumour: proteomics in cancer research. *Clin Sci (Lond)* 108 (5), 369-383 (2005).

[189] Silvestri E, Lombardi A, de Lange P, Glinni D, Senese R, Cioffi F, Lanni A, Goglia F, Moreno M. Studies of complex biological systems with applications to molecular medicine: the need to integrate transcriptomic and proteomic approaches. *J Biomed Biotechnol* 2011, 810242.

[190] Vlahou, A. & Fountoulakis, M., Proteomic approaches in the search for disease biomarkers. *J Chromatogr B Analyt Technol Biomed Life Sci* 814 (1), 11-19 (2005).

[191] Rennert, G., Are we getting closer to molecular population screening for colorectal cancer? *J Natl Cancer Inst* 101 (13), 902-903 (2009).

[192] O'Connell, J.B., Maggard, M.A., & Ko, C.Y., Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst* 96 (19), 1420-1425 (2004).

[193] Schubert, W., Topological proteomics, toponomics, MELK-technology. *Adv Biochem Eng Biotechnol* 83, 189-209 (2003).

[194] Schubert, W., Exploring molecular networks directly in the cell. *Cytometry A* 69 (3), 109-112 (2006).

[195] Schubert, W., A three-symbol code for organized proteomes based on cyclical imaging of protein locations. *Cytometry A* 71 (6), 352-360 (2007).

[196] Vogelstein, B. & Kinzler, K.W., Cancer genes and the pathways they control. *Nat Med* 10 (8), 789-799 (2004).

[197] Dietmaier W, Wallinger S, Bocker T, Kullmann F, Fishel R, Rüschoff J. Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. *Cancer Res* 57 (21), 4749-4756 (1997).

[198] Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, Cavalieri RJ, Boland CR. Genetics, natural history, tumor spectrum, and pathology of

hereditary nonpolyposis colorectal cancer: an updated review. *Gastroenterology* 104 (5), 1535-1549 (1993).

[199] Ilyas, M., Straub, J., Tomlinson, I.P., & Bodmer, W.F., Genetic pathways in colorectal and other cancers. *Eur J Cancer* 35 (3), 335-351 (1999).

[200] Popat, S., Hubner, R., & Houlston, R.S., Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* 23 (3), 609-618 (2005).

[201] Des Guetz G, Schischmanoff O, Nicolas P, Perret GY, Morere JF, Uzzan B. Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis. *Eur J Cancer* 45 (10), 1890-1896 (2009).

[202] Suter, C.M., Martin, D.I., & Ward, R.L., Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 36 (5), 497-501 (2004).

[203] Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF, Tsui WY, Lo MW, Tam WY, Li VS, Leung SY. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat Genet* 38 (10), 1178-1183 (2006).

[204] Gazzoli, I., Loda, M., Garber, J., Syngal, S., & Kolodner, R.D., A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res* 62 (14), 3925-3928 (2002).

[205] Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ, Tsui WY, Kong CK, Brunner HG, van Kessel AG, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N. Heritable somatic methylation and inactivation of MSH2 in

families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 41 (1), 112-117 (2009).

206 Kovacs, M.E., Papp, J., Szentirmay, Z., Otto, S., & Olah, E., Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum Mutat* 30 (2), 197-203 (2009).

207 Easton, D.F. & Eeles, R.A., Genome-wide association studies in cancer. *Hum Mol Genet* 17 (R2), R109-115 (2008).

208 Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K; CORGI Consortium, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB, Houlston R. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39 (8), 984-988 (2007).

209 Goel, A. & Boland, C.R., Recent insights into the pathogenesis of colorectal cancer. *Curr Opin Gastroenterol* 26 (1), 47-52.

210 Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Björklund M, Wei G, Yan J, Niittymäki I, Mecklin JP, Järvinen H, Ristimäki A, Di-Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Palin K, Ukkonen E, Karhu A, Taipale J, Aaltonen LA. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41 (8), 885-890 (2009).

211 Howe, J.R. & Guillem, J.G., The genetics of colorectal cancer. *Surg Clin North Am* 77 (1), 175-195 (1997).

[212] Cerottini, J.P., Caplin, S., Saraga, E., Givel, J.C., & Benhattar, J., The type of K-ras mutation determines prognosis in colorectal cancer. *Am J Surg* 175 (3), 198-202 (1998).

[213] Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P, Vogelstein B. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* 256 (5053), 102-105 (1992).

[214] Fearnhead, N.S., Britton, M.P., & Bodmer, W.F., The ABC of APC. *Hum Mol Genet* 10 (7), 721-733 (2001).

[215] Halling KC, French AJ, McDonnell SK, Burgart LJ, Schaid DJ, Peterson BJ, Moon-Tasson L, Mahoney MR, Sargent DJ, O'Connell MJ, Witzig TE, Farr GH Jr, Goldberg RM, Thibodeau SN. Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers. *J Natl Cancer Inst* 91 (15), 1295-1303 (1999).

[216] Vousden, K.H., p53: death star. *Cell* 103 (5), 691-694 (2000).

[217] Kressner U, Inganäs M, Byding S, Blikstad I, Påhlman L, Glimelius B, Lindmark G. Prognostic value of p53 genetic changes in colorectal cancer. *J Clin Oncol* 17 (2), 593-599 (1999).

[218] Ahnen DJ, Feigl P, Quan G, Fenoglio-Preiser C, Lovato LC, Bunn PA Jr, Stemmerman G, Wells JD, Macdonald JS, Meyskens FL Jr. Ki-ras mutation and p53 overexpression predict the clinical behavior of colorectal cancer: a Southwest Oncology Group study. *Cancer Res* 58 (6), 1149-1158 (1998).

[219] Watanabe T, Wu TT, Catalano PJ, Ueki T, Satriano R, Haller DG, Benson AB 3rd, Hamilton SR. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med* 344 (16), 1196-1206 (2001).

[220] Zhu, J. & Yao, X., Use of DNA methylation for cancer detection and molecular classification. *J Biochem Mol Biol* 40 (2), 135-141 (2007).

[221] Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M, Song X, Lesche R, Liebenberg V, Ebert M, Molnar B, Grützmann R, Pilarsky C, Sledziewski A. DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin Chem* 54 (2), 414-423 (2008).

[222] Wang, Z., Gerstein, M., & Snyder, M., RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10 (1), 57-63 (2009).

[223] Clark, T.A., Sugnet, C.W., & Ares, M., Jr., Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296 (5569), 907-910 (2002).

[224] Boguski, M.S., Tolstoshev, C.M., & Bassett, D.E., Jr., Gene discovery in dbEST. *Science* 265 (5181), 1993-1994 (1994).

[225] Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, Guyer M, Peck AM, Derge JG, Lipman D, Collins FS, Jang W, Sherry S, Feolo M, Misquitta L, Lee E, Rotmistrovsky K, Greenhut SF, *et al*. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res* 14 (10B), 2121-2127 (2004).

[226] Harbers, M. & Carninci, P., Tag-based approaches for transcriptome research and genome annotation. *Nat Methods* 2 (7), 495-502 (2005).

[227] Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M, Hayashizaki Y, Carninci P. CAGE: cap analysis of gene expression. *Nat Methods* 3 (3), 211-222 (2006).

[228] Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18 (6), 630-634 (2000).

[229] Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, Kramer M, Woychik R. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Brief Funct Genomic Proteomic* 1 (1), 95-104 (2002).

[230] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320 (5881), 1344-1349 (2008).

[231] Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., & Gilad, Y., RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18 (9), 1509-1517 (2008).

[232] Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453 (7199), 1239-1243 (2008).

[233] Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., & Wold, B., Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5 (7), 621-628 (2008).

[234] Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133 (3), 523-536 (2008).

[235] Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5 (7), 613-619 (2008).

[236] Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* 17 (7), 1636-1647 (2008).

[237] Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, McDonald H, Varhol R, Jones S, Marra M. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45 (1), 81-94 (2008).

[238] Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, Perou CM. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7, 96 (2006).

[239] Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98 (4), 262-272 (2006).

[240] Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM, West M, Nevins JR, Huang AT. Gene expression predictors of breast cancer outcomes. *Lancet* 361 (9369), 1590-1596 (2003).

[241] van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415 (6871), 530-536 (2002).

[242] van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347 (25), 1999-2009 (2002).

[243] Lengauer, C., Kinzler, K.W., & Vogelstein, B., Genetic instabilities in human cancers. *Nature* 396 (6712), 643-649 (1998).

[244] Saccani Jotti G, Fontanesi M, Orsi N, Sarli L, Pietra N, Peracchia A, Sansebastiano G, Becchi G. DNA content in human colon cancer and non-neoplastic adjacent mucosa. *Int J Biol Markers* 10 (1), 11-16 (1995).

[245] Habermann J, Lenander C, Roblick UJ, Krüger S, Ludwig D, Alaiya A, Freitag S, Dümbgen L, Bruch HP, Stange E, Salo S, Tryggvason K, Auer G, Schimmelpenning H. Ulcerative colitis and colorectal carcinoma: DNA-profile, laminin-5 gamma2 chain and cyclin A expression as early markers for risk assessment. *Scand J Gastroenterol* 36 (7), 751-758 (2001).

[246] Chen, H.S., Sheen-Chen, S.M., & Lu, C.C., DNA index and S-phase fraction in curative resection of colorectal adenocarcinoma: analysis of prognosis and current trends. *World J Surg* 26 (5), 626-630 (2002).

[247] Habermann JK, Paulsen U, Roblick UJ, Upender MB, McShane LM, Korn EL, Wangsa D, Krüger S, Duchrow M, Bruch HP, Auer G, Ried T. Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis. *Genes Chromosomes Cancer* 46 (1), 10-26 (2007).

[248] Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* 93 (20), 10614-10619 (1996).

[249] Shih, W., Chetty, R., & Tsao, M.S., Expression profiling by microarrays in colorectal cancer (Review). *Oncol Rep* 13 (3), 517-524 (2005).

[250] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. *Nature* 435 (7043), 834-838 (2005).

[251] Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, Barad O, Bentwich Z, Szafranska AE, Labourier E, Raymond CK, Roberts BS, Juhl H, Kinzler KW, Vogelstein B, Velculescu VE. The colorectal microRNAome. *Proc Natl Acad Sci U S A* 103 (10), 3687-3692 (2006).

[252] Monzo M, Navarro A, Bandres E, Artells R, Moreno I, Gel B, Ibeas R, Moreno J, Martinez F, Diaz T, Martinez A, Balagué O, Garcia-Foncillas J. Overlapping expression of microRNAs in human embryonic colon and colorectal cancer. *Cell Res* 18 (8), 823-833 (2008).

[253] Bandrés E, Cubedo E, Agirre X, Malumbres R, Zárate R, Ramirez N, Abajo A, Navarro A, Moreno I, Monzó M, García-Foncillas J. Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol Cancer* 5, 29 (2006).

[254] Bosch Roig CE, Roselló-Sastre E, Alonso Hernández S, Almenar Cubells D, Grau Cardona E, Camarasa Lillo N, Bautista D, Molins Palau C. Prognostic value of the detection of lymph node micrometastases in colon cancer. *Clin Transl Oncol* 10 (9), 572-578 (2008).

[255] Misek, D.E., Imafuku, Y., & Hanash, S.M., Application of proteomic technologies to tumor analysis. *Pharmacogenomics* 5 (8), 1129-1137 (2004).

[256] Everley, P.A. & Zetter, B.R., Proteomics in tumor progression and metastasis. *Ann N Y Acad Sci* 1059, 1-10 (2005).

[257] O'Farrell, P.H., High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250 (10), 4007-4021 (1975).

[258] Kleparnik, K. & Bocek, P., Electrophoresis today and tomorrow: Helping biologists' dreams come true. *Bioessays* 32 (3), 218-226.

[259] Gorg, A., Weiss, W., & Dunn, M.J., Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4 (12), 3665-3685 (2004).

[260] Cho, W.C. & Cheng, C.H., Oncoproteomics: current trends and future perspectives. *Expert Rev Proteomics* 4 (3), 401-410 (2007).

[261] Martín B, Aragüés R, Sanz R, Oliva B, Boluda S, Martínez A, Sierra A. Biological pathways contributing to organ-specific phenotype of brain metastatic cells. *J Proteome Res* 7 (3), 908-920 (2008).

[262] Hatakeyama H, Kondo T, Fujii K, Nakanishi Y, Kato H, Fukuda S, Hirohashi S. Protein clusters associated with carcinogenesis, histological differentiation and nodal metastasis in esophageal cancer. *Proteomics* 6 (23), 6300-6316 (2006).

[263] Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I. A novel experimental design for comparative two-dimensional gel analysis: two-

dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 3 (1), 36-44 (2003).

[264] Unlu, M., Morgan, M.E., & Minden, J.S., Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18 (11), 2071-2077 (1997).

[265] Han, X., Aslanian, A., & Yates, J.R., 3rd, Mass spectrometry for proteomics. *Curr Opin Chem Biol* 12 (5), 483-490 (2008).

[266] Pan S, Aebersold R, Chen R, Rush J, Goodlett DR, McIntosh MW, Zhang J, Brentnall TA. Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res* 8 (2), 787-797 (2009).

[267] McAlister, G.C., Phanstiel, D., Wenger, C.D., Lee, M.V., & Coon, J.J., Analysis of tandem mass spectra by FTMS for improved large-scale proteomics with superior protein quantification. *Anal Chem* 82 (1), 316-322.

[268] Kicman, A.T., Parkin, M.C., & Iles, R.K., An introduction to mass spectrometry based proteomics-detection and characterization of gonadotropins and related molecules. *Mol Cell Endocrinol* 260-262, 212-227 (2007).

[269] Liao CC, Mehta A, Ward NJ, Marsh S, Arulampalam T, Norton JD. Analysis of post-operative changes in serum protein expression profiles from colorectal cancer patients by MALDI-TOF mass spectrometry: a pilot methodological study. *World J Surg Oncol* 8, 33.

[270] Wei J, Xu G, Wu M, Zhang Y, Li Q, Liu P, Zhu T, Song A, Zhao L, Han Z, Chen G, Wang S, Meng L, Zhou J, Lu Y, Wang S, Ma D. Overexpression of vimentin contributes to prostate cancer invasion and metastasis via src regulation. *Anticancer Res* 28 (1A), 327-334 (2008).

[271] Schwartz SA, Weil RJ, Thompson RC, Shyr Y, Moore JH, Toms SA, Johnson MD, Caprioli RM. Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. *Cancer Res* 65 (17), 7674-7681 (2005).

[272] Vydra J, Selicharová I, Smutná K, Sanda M, Matousková E, Bursíková E, Prchalová M, Velenská Z, Coufal D, Jirácek J., Two-dimensional electrophoretic comparison of metastatic and non-metastatic human breast tumors using in vitro cultured epithelial cells derived from the cancer tissues. *BMC Cancer* 8, 107 (2008).

[273] Tang, N., Tornatore, P., & Weinberger, S.R., Current developments in SELDI affinity technology. *Mass Spectrom Rev* 23 (1), 34-44 (2004).

[274] Jr GW, Cazares LH, Leung SM, Nasim S, Adam BL, Yip TT, Schellhammer PF, Gong L, Vlahou A. Proteinchip(R) surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures. *Prostate Cancer Prostatic Dis* 2 (5/6), 264-276 (1999).

[275] Conrads, T.P., Hood, B.L., Issaq, H.J., & Veenstra, T.D., Proteomic patterns as a diagnostic tool for early-stage cancer: a review of its progress to a clinically relevant tool. *Mol Diagn* 8 (2), 77-85 (2004).

[276] De Bock M, de Seny D, Meuwis MA, Chapelle JP, Louis E, Malaise M, Merville MP, Fillet M. Challenges for biomarker discovery in body fluids using SELDI-TOF-MS. *J Biomed Biotechnol* 2010, 906082.

[277] Wang Q, Shen J, Li ZF, Jie JZ, Wang WY, Wang J, Zhang ZT, Li ZX, Yan L, Gu J. Limitations in SELDI-TOF MS whole serum proteomic profiling with IMAC surface to specifically detect colorectal cancer. *BMC Cancer* 9, 287 (2009).

[278] Smith FM, Gallagher WM, Fox E, Stephens RB, Rexhepaj E, Petricoin EF 3rd, Liotta L, Kennedy MJ, Reynolds JV. Combination of SELDI-TOF-MS and data mining provides early-stage response prediction for rectal tumors undergoing multimodal neoadjuvant therapy. *Ann Surg* 245 (2), 259-266 (2007).

[279] Ciordia, S., de Los Rios, V., & Albar, J.P., Contributions of advanced proteomics technologies to cancer diagnosis. *Clin Transl Oncol* 8 (8), 566-580 (2006).

[280] Qiu FM, Yu JK, Chen YD, Jin QF, Sui MH, Huang J. Mining novel biomarkers for prognosis of gastric cancer with serum proteomics. *J Exp Clin Cancer Res* 28, 126 (2009).

[281] Asara, J.M., Christofk, H.R., Freimark, L.M., & Cantley, L.C., A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* 8 (5), 994-999 (2008).

[282] Faça VM, Ventura AP, Fitzgibbon MP, Pereira-Faça SR, Pitteri SJ, Green AE, Ireton RC, Zhang Q, Wang H, O'Briant KC, Drescher CW, Schummer M, McIntosh MW, Knudsen BS, Hanash SM. Proteomic analysis of ovarian cancer cells reveals dynamic processes of protein secretion and shedding of extra-cellular domains. *PLoS ONE* 3 (6), e2425 (2008).

[283] Zhang DY, Ye F, Gao L, Liu X, Zhao X, Che Y, Wang H, Wang L, Wu J, Song D, Liu W, Xu H, Jiang B, Zhang W, Wang J, Lee P. Proteomics, pathway array and signaling network-based medicine in cancer. *Cell Div* 4, 20 (2009).

[284] Bertone, P. & Snyder, M., Advances in functional protein microarray technology. *FEBS J* 272 (21), 5400-5411 (2005).

[285] Harmsen, M.M. & De Haard, H.J., Properties, production, and applications of camelid single-domain antibody fragments. *Appl Microbiol Biotechnol* 77 (1), 13-22 (2007).

[286] Kastelic D, Frković-Grazio S, Baty D, Truan G, Komel R, Pompon D. A single-step procedure of recombinant library construction for the selection of efficiently produced llama VH binders directed against cancer markers. *J Immunol Methods* 350 (1-2), 54-62 (2009).

[287] Osterfeld SJ, Yu H, Gaster RS, Caramuta S, Xu L, Han SJ, Hall DA, Wilson RJ, Sun S, White RL, Davis RW, Pourmand N, Wang SX. Multiplex protein assays based on real-time magnetic nanotag sensing. *Proc Natl Acad Sci U S A* 105 (52), 20637-20640 (2008).

[288] Hall, D.A., Ptacek, J., & Snyder, M., Protein microarray technology. *Mech Ageing Dev* 128 (1), 161-167 (2007).

[289] Fei YY, Landry JP, Sun YS, Zhu XD, Luo JT, Wang XB, Lam KS. A novel high-throughput scanning microscope for label-free detection of protein and small-molecule chemical microarrays. *Rev Sci Instrum* 79 (1), 013708 (2008).

[290] Tahara H, Sato M, Thurin M, Wang E, Butterfield LH, Disis ML, Fox BA, Lee PP, Khleif SN, Wigginton JM, Ambs S, Akutsu Y, Chaussabel D, Doki Y, Eremin O, Fridman WH, Hirohashi Y, Imai K, Jacobson J, Jinushi M, Kanamoto A, Kashani-Sabet M, Kato K, Kawakami Y, Kirkwood JM, Kleen TO, Lehmann PV, Liotta L, Lotze MT, *et al.*, Emerging concepts in biomarker discovery; the US-Japan Workshop on Immunological Molecular Markers in Oncology. *J Transl Med* 7, 45 (2009).

[291] Grubb RL, Deng J, Pinto PA, Mohler JL, Chinnaiyan A, Rubin M, Linehan WM, Liotta LA, Petricoin EF, Wulfkuhle JD. Pathway biomarker profiling of localized and metastatic human prostate cancer reveal metastatic and prognostic signatures. *J Proteome Res* 8 (6), 3044-3054 (2009).

[292] Wulfkuhle JD, Speer R, Pierobon M, Laird J, Espina V, Deng J, Mammano E, Yang SX, Swain SM, Nitti D, Esserman LJ, Belluco C, Liotta LA, Petricoin EF 3rd. Multiplexed cell signaling analysis of human breast cancer applications for personalized therapy. *J Proteome Res* 7 (4), 1508-1517 (2008).

[293] Gulmann C, Sheehan KM, Conroy RM, Wulfkuhle JD, Espina V, Mullarkey MJ, Kay EW, Liotta LA, Petricoin EF 3rd. Quantitative cell signalling analysis reveals down-regulation of MAPK pathway activation in colorectal cancer. *J Pathol* 218 (4), 514-519 (2009).

[294] He, M. & Taussig, M.J., Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acids Res* 29 (15), E73-73 (2001).

[295] He, M. & Taussig, M.J., Production of protein arrays by cell-free systems. *Methods Mol Biol* 484, 207-215 (2008).

[296] Sanchez-Carbayo, M., Antibody array-based technologies for cancer protein profiling and functional proteomic analyses using serum and tissue specimens. *Tumour Biol* 31 (2), 103-112.

[297] Wulfkuhle JD, Aquino JA, Calvert VS, Fishman DA, Coukos G, Liotta LA, Petricoin EF 3rd. Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics* 3 (11), 2085-2090 (2003).

[298] Nettikadan S, Radke K, Johnson J, Xu J, Lynch M, Mosher C, Henderson E. Detection and quantification of protein biomarkers from fewer than 10 cells. *Mol Cell Proteomics* 5 (5), 895-901 (2006).

[299] Song S, Li B, Wang L, Wu H, Hu J, Li M, Fan C. A cancer protein microarray platform using antibody fragments and its clinical applications. *Mol Biosyst* 3 (2), 151-158 (2007).

[300] Craven, R.A. & Banks, R.E., Laser capture microdissection and proteomics: possibilities and limitation. *Proteomics* 1 (10), 1200-1204 (2001).

[301] Tan, N.C., Fisher, W.G., Rosenblatt, K.P., & Garner, H.R., Application of multiple statistical tests to enhance mass spectrometry-based biomarker discovery. *BMC Bioinformatics* 10, 144 (2009).

[302] Blueggel, M., Chamrad, D., & Meyer, H.E., Bioinformatics in proteomics. *Curr Pharm Biotechnol* 5 (1), 79-88 (2004).

[303] Jungblut PR, Zimny-Arndt U, Zeindl-Eberhart E, Stulik J, Koupilova K, Pleissner KP, Otto A, Müller EC, Sokolowska-Köhler W, Grabher G, Stöffler G. Proteomics in human disease: cancer, heart and infectious diseases. *Electrophoresis* 20 (10), 2100-2110 (1999).

[304] Friedman DB, Hill S, Keller JW, Merchant NB, Levy SE, Coffey RJ, Caprioli RM. Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 4 (3), 793-811 (2004).

[305] Alfonso P, Núñez A, Madoz-Gurpide J, Lombardia L, Sánchez L, Casal JI. Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis. *Proteomics* 5 (10), 2602-2611 (2005).

[306] Allal AS, Kähne T, Reverdin AK, Lippert H, Schlegel W, Reymond MA. Radioresistance-related proteins in rectal cancer. *Proteomics* 4 (8), 2261-2269 (2004).

[307] Kwong KY, Bloom GC, Yang I, Boulware D, Coppola D, Haseman J, Chen E, McGrath A, Makusky AJ, Taylor J, Steiner S, Zhou J, Yeatman TJ, Quackenbush

J. Synchronous global assessment of gene and protein expression in colorectal cancer progression. *Genomics* 86 (2), 142-158 (2005).

[308] Nishizuka S, Chen ST, Gwadry FG, Alexander J, Major SM, Scherf U, Reinhold WC, Waltham M, Charboneau L, Young L, Bussey KJ, Kim S, Lababidi S, Lee JK, Pittaluga S, Scudiero DA, Sausville EA, Munson PJ, Petricoin EF 3rd, Liotta LA, Hewitt SM, Raffeld M, Weinstein JN. Diagnostic markers that distinguish colon and ovarian adenocarcinomas: identification by genomic, proteomic, and tissue array profiling. *Cancer Res* 63 (17), 5243-5250 (2003).

[309] Roblick UJ, Bader FG, Lenander C, Hellman U, Zimmermann K, Becker S, Ost A, Alaiya A, Bruch HP, Keller R, Mirow L, Franzén B, Ried T, Auer G, Habermann JK. Undifferentiated pelvic adenocarcinomas: diagnostic potential of protein profiling and multivariate analysis. *Int J Colorectal Dis* 23 (5), 483-491 (2008).

[310] Wang Y, Jatkoe T, Zhang Y, Mutch MG, Talantov D, Jiang J, McLeod HL, Atkins D. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol* 22 (9), 1564-1571 (2004).

[311] Feldman, A.L., Espina, V., Petricoin, E.F., 3rd, Liotta, L.A., & Rosenblatt, K.P., Use of proteomic patterns to screen for gastrointestinal malignancies. *Surgery* 135 (3), 243-247 (2004).

[312] Thadikkaran L, Siegenthaler MA, Crettaz D, Queloz PA, Schneider P, Tissot JD. Recent advances in blood-related proteomics. *Proteomics* 5 (12), 3019-3034 (2005).

[313] Gold, P. & Freedman, S.O., Specific carcinoembryonic antigens of the human digestive system. *J Exp Med* 122 (3), 467-481 (1965).

[314] Slentz, K., Senagore, A., Hibbert, J., Mazier, W.P., & Talbott, T.M., Can preoperative and postoperative CEA predict survival after colon cancer resection? *Am Surg* 60 (7), 528-531; discussion 531-522 (1994).

[315] Chapman, M.A., Buckley, D., Henson, D.B., & Armitage, N.C., Preoperative carcinoembryonic antigen is related to tumour stage and long-term survival in colorectal cancer. *Br J Cancer* 78 (10), 1346-1349 (1998).

[316] Harrison, L.E., Guillem, J.G., Paty, P., & Cohen, A.M., Preoperative carcinoembryonic antigen predicts outcomes in node-negative colon cancer patients: a multivariate analysis of 572 patients. *J Am Coll Surg* 185 (1), 55-59 (1997).

[317] Fletcher, R.H., Carcinoembryonic antigen. *Ann Intern Med* 104 (1), 66-73 (1986).

[318] Glover C, Douse P, Kane P, Karani J, Meire H, Mohammadtaghi S, Allen-Mersh TG. Accuracy of investigations for asymptomatic colorectal liver metastases. *Dis Colon Rectum* 45 (4), 476-484 (2002).

[319] Mori M, Tomoda H, Ishida T, Kido A, Shimono R, Matsushima T, Kuwano H, Sugimachi K. Surgical resection of pulmonary metastases from colorectal adenocarcinoma. Special reference to repeated pulmonary resections. *Arch Surg* 126 (10), 1297-1301; discussion 1302 (1991).

[320] Duffy, M.J., Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? *Clin Chem* 47 (4), 624-630 (2001).

[321] Duffy MJ, van Dalen A, Haglund C, Hansson L, Klapdor R, Lamerz R, Nilsson O, Sturgeon C, Topolcan O. Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. *Eur J Cancer* 39 (6), 718-727 (2003).

[322] Hsu, T.C., Unusual elevation of CEA in a patient with history of colon cancer. *Jpn J Clin Oncol* 36 (12), 811-813 (2006).

[323] Whitaker, D. & Shilkin, K.B., Carcinoembryonic antigen in tissue diagnosis of malignant mesothelioma. *Lancet* 1 (8234), 1369 (1981).

[324] Lee H, Rhee H, Kang HJ, Kim HS, Min BS, Kim NK, Kim H. Macrophage migration inhibitory factor may be used as an early diagnostic marker in colorectal carcinomas. *Am J Clin Pathol* 129 (5), 772-779 (2008).

[325] Duffy, M.J., CA 19-9 as a marker for gastrointestinal cancers: a review. *Ann Clin Biochem* 35 (Pt 3), 364-370 (1998).

[326] Magnani JL, Nilsson B, Brockhaus M, Zopf D, Steplewski Z, Koprowski H, Ginsburg V. A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. *J Biol Chem* 257 (23), 14365-14369 (1982).

[327] Hundt, S., Haug, U., & Brenner, H., Blood markers for early detection of colorectal cancer: a systematic review. *Cancer Epidemiol Biomarkers Prev* 16 (10), 1935-1953 (2007).

[328] Maurel J, Nadal C, Garcia-Albeniz X, Gallego R, Carcereny E, Almendro V, Mármol M, Gallardo E, Maria Augé J, Longarón R, Martínez-Fernandez A, Molina R, Castells A, Gascón P. Serum matrix metalloproteinase 7 levels identifies poor prognosis advanced colorectal cancer patients. *Int J Cancer* 121 (5), 1066-1071 (2007).

[329] Hurst NG, Stocken DD, Wilson S, Keh C, Wakelam MJ, Ismail T. Elevated serum matrix metalloproteinase 9 (MMP-9) concentration predicts the presence of colorectal neoplasia in symptomatic patients. *Br J Cancer* 97 (7), 971-977 (2007).

[330] Soroush, A.R., Zadeh, H.M., Moemeni, M., Shakiba, B., & Elmi, S., Plasma prolactin in patients with colorectal cancer. *BMC Cancer* 4, 97 (2004).

[331] Mroczko B, Groblewska M, Wereszczynska-Siemiatkowska U, Kedra B, Konopko M, Szmitkowski M. The diagnostic value of G-CSF measurement in the sera of colorectal cancer and adenoma patients. *Clin Chim Acta* 371 (1-2), 143-147 (2006).

[332] Mroczko B, Groblewska M, Wereszczyńska-Siemiatkowska U, Okulczyk B, Kedra B, Łaszewicz W, Dabrowski A, Szmitkowski M. Serum macrophage-colony stimulating factor levels in colorectal cancer patients correlate with lymph node metastasis and poor prognosis. *Clin Chim Acta* 380 (1-2), 208-212 (2007).

[333] Saito, N. & Kameoka, S., Serum laminin is an independent prognostic factor in colorectal cancer. *Int J Colorectal Dis* 20 (3), 238-244 (2005).

[334] Holten-Andersen MN, Murphy G, Nielsen HJ, Pedersen AN, Christensen IJ, Høyer-Hansen G, Brünner N, Stephens RW. Quantitation of TIMP-1 in plasma of healthy blood donors and patients with advanced cancer. *Br J Cancer* 80 (3-4), 495-503 (1999).

[335] Holten-Andersen MN, Fenger C, Nielsen HJ, Rasmussen AS, Christensen IJ, Brünner N, Kronborg O. Plasma TIMP-1 in patients with colorectal adenomas: a prospective study. *Eur J Cancer* 40 (14), 2159-2164 (2004).

[336] Sørensen NM, Schrohl AS, Jensen V, Christensen IJ, Nielsen HJ, Brünner N. Comparative studies of tissue inhibitor of metalloproteinases-1 in plasma, serum and tumour tissue extracts from patients with primary colorectal cancer. *Scand J Gastroenterol* 43 (2), 186-191 (2008).

[337] Holten-Andersen, M. *et al.*, Association between preoperative plasma levels of tissue inhibitor of metalloproteinases 1 and rectal cancer patient survival. a validation study. *Eur J Cancer* 40 (1), 64-72 (2004).

[338] Holten-Andersen MN, Stephens RW, Nielsen HJ, Murphy G, Christensen IJ, Stetler-Stevenson W, Brünner N. High preoperative plasma tissue inhibitor of metalloproteinase-1 levels are associated with short survival of patients with colorectal cancer. *Clin Cancer Res* 6 (11), 4292-4299 (2000).

[339] Offenberg, H., Brunner, N., Mansilla, F., Orntoft Torben, F., & Birkenkamp-Demtroder, K., TIMP-1 expression in human colorectal cancer is associated with TGF-B1, LOXL2, INHBA1, TNF-AIP6 and TIMP-2 transcript profiles. *Mol Oncol* 2 (3), 233-240 (2008).

[340] Wang, C.S., Wu, T.L., Tsao, K.C., & Sun, C.F., Serum TIMP-1 in gastric cancer patients: a potential prognostic biomarker. *Ann Clin Lab Sci* 36 (1), 23-30 (2006).

[341] Naruo S, Kanayama H, Takigawa H, Kagawa S, Yamashita K, Hayakawa T. Serum levels of a tissue inhibitor of metalloproteinases-1 (TIMP-1) in bladder cancer patients. *Int J Urol* 1 (3), 228-231 (1994).

[342] Ylisirniö S, Höyhtyä M, Mäkitaro R, Pääakkö P, Risteli J, Kinnula VL, Turpeenniemi-Hujanen T, Jukkola A. Elevated serum levels of type I collagen degradation marker ICTP and tissue inhibitor of metalloproteinase (TIMP) 1 are associated with poor prognosis in lung cancer. *Clin Cancer Res* 7 (6), 1633-1637 (2001).

[343] Manenti L, Paganoni P, Floriani I, Landoni F, Torri V, Buda A, Taraboletti G, Labianca R, Belotti D, Giavazzi R. Expression levels of vascular endothelial growth factor, matrix metalloproteinases 2 and 9 and tissue inhibitor of

metalloproteinases 1 and 2 in the plasma of patients with ovarian carcinoma. *Eur J Cancer* 39 (13), 1948-1956 (2003).

344 Roessler M, Rollinger W, Palme S, Hagmann ML, Berndt P, Engel AM, Schneidinger B, Pfeffer M, Andres H, Karl J, Bodenmüller H, Rüschoff J, Henkel T, Rohr G, Rossol S, Rösch W, Langen H, Zolg W, Tacke Identification of nicotinamide N-methyltransferase as a novel serum tumor marker for colorectal cancer. *Clin Cancer Res* 11 (18), 6550-6557 (2005).

345 Roessler M, Rollinger W, Mantovani-Endl L, Hagmann ML, Palme S, Berndt P, Engel AM, Pfeffer M, Karl J, Bodenmüller H, Rüschoff J, Henkel T, Rohr G, Rossol S, Rösch W, Langen H, Zolg W, Tacke M. Identification of PSME3 as a novel serum tumor marker for colorectal cancer by combining two-dimensional polyacrylamide gel electrophoresis with a strictly mass spectrometry-based approach for data analysis. *Mol Cell Proteomics* 5 (11), 2092-2101 (2006).

346 Wu CC, Chen HC, Chen SJ, Liu HP, Hsieh YY, Yu CJ, Tang R, Hsieh LL, Yu JS, Chang YS. Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics* 8 (2), 316-332 (2008).

347 Albrethsen J, Bøgebo R, Gammeltoft S, Olsen J, Winther B, Raskov H. Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1-3) in colon cancer serum and tumours: a biomarker study. *BMC Cancer* 5, 8 (2005).

348 Melle C, Ernst G, Schimmel B, Bleul A, Thieme H, Kaufmann R, Mothes H, Settmacher U, Claussen U, Halbhuber KJ, Von Eggeling F. Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer. *Gastroenterology* 129 (1), 66-73 (2005).

[349] Ward DG, Suggett N, Cheng Y, Wei W, Johnson H, Billingham LJ, Ismail T, Wakelam MJ, Johnson PJ, Martin A. Identification of serum biomarkers for colon cancer by proteomic analysis. *Br J Cancer* 94 (12), 1898-1905 (2006).

[350] Habermann JK, Roblick UJ, Luke BT, Prieto DA, Finlay WJ, Podust VN, Roman JM, Oevermann E, Schiedeck T, Homann N, Duchrow M, Conrads TP, Veenstra TD, Burt SK, Bruch HP, Auer G, Ried T. Increased serum levels of complement C3a anaphylatoxin indicate the presence of colorectal tumors. *Gastroenterology* 131 (4), 1020-1029; quiz 1284 (2006).

[351] Yu, J.K., Chen, Y.D., & Zheng, S., An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J Gastroenterol* 10 (21), 3127-3131 (2004).

[352] Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, Ederer F. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med* 328 (19), 1365-1371 (1993).

[353] Schnoll-Sussman, F., Markowitz, A.J., & Winawer, S.J., Screening and surveillance for colorectal cancer. *Semin Oncol* 27 (5 Suppl 10), 10-21 (2000).

[354] Woolf, S.H., A smarter strategy? Reflections on fecal DNA screening for colorectal cancer. *N Engl J Med* 351 (26), 2755-2758 (2004).

[355] Atkin, W. & Martin, J.P., Stool DNA-based colorectal cancer detection: finding the needle in the haystack. *J Natl Cancer Inst* 93 (11), 798-799 (2001).

[356] Boynton, K.A., Summerhayes, I.C., Ahlquist, D.A., & Shuber, A.P., DNA integrity as a potential marker for stool-based detection of colorectal cancer. *Clin Chem* 49 (7), 1058-1065 (2003).

[357] Dong SM, Traverso G, Johnson C, Geng L, Favis R, Boynton K, Hibi K, Goodman SN, D'Allessio M, Paty P, Hamilton SR, Sidransky D, Barany F, Levin B, Shuber A, Kinzler KW, Vogelstein B, Jen J. Detecting colorectal cancer in stool with the use of multiple genetic targets. *J Natl Cancer Inst* 93 (11), 858-865 (2001).

[358] Imperiale, T.F., Ransohoff, D.F., Itzkowitz, S.H., Turnbull, B.A., & Ross, M.E., Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 351 (26), 2704-2714 (2004).

[359] Aadland, E. & Fagerhol, M.K., Faecal calprotectin: a marker of inflammation throughout the intestinal tract. *Eur J Gastroenterol Hepatol* 14 (8), 823-825 (2002).

[360] Summerton, C.B., Longlands, M.G., Wiener, K., & Shreeve, D.R., Faecal calprotectin: a marker of inflammation throughout the intestinal tract. *Eur J Gastroenterol Hepatol* 14 (8), 841-845 (2002).

[361] Limburg PJ, Devens ME, Harrington JJ, Diehl NN, Mahoney DW, Ahlquist DA. Prospective evaluation of fecal calprotectin as a screening biomarker for colorectal neoplasia. *Am J Gastroenterol* 98 (10), 2299-2305 (2003).

[362] Kim, H.J., Yu, M.H., Kim, H., Byun, J., & Lee, C., Noninvasive molecular biomarkers for the detection of colorectal cancer. *BMB Rep* 41 (10), 685-692 (2008).

[363] Schneider, J., Bitterlich, N., & Schulze, G., Improved sensitivity in the diagnosis of gastro-intestinal tumors by fuzzy logic-based tumor marker profiles including the tumor M2-PK. *Anticancer Res* 25 (3A), 1507-1515 (2005).

[364] Zhang, B., Chen, J.Y., Chen, D.D., Wang, G.B., & Shen, P., Tumor type M2 pyruvate kinase expression in gastric cancer, colorectal cancer and controls. *World J Gastroenterol* 10 (11), 1643-1646 (2004).

[365] Han M, Liew CT, Zhang HW, Chao S, Zheng R, Yip KT, Song ZY, Li HM, Geng XP, Zhu LX, Lin JJ, Marshall KW, Liew CC. Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. *Clin Cancer Res* 14 (2), 455-460 (2008).

[366] Minamoto, T., Mai, M., & Ronai, Z., K-ras mutation: early detection in molecular diagnosis and risk assessment of colorectal, pancreas, and lung cancers--a review. *Cancer Detect Prev* 24 (1), 1-12 (2000).

[367] Ahlquist, D.A. & Shuber, A.P., Stool screening for colorectal cancer: evolution from occult blood to molecular markers. *Clin Chim Acta* 315 (1-2), 157-168 (2002).

[368] Capturing cancer stem cells. *Nat Med* 14 (8), 814 (2008).

[369] Patel BB, Yu Y, Du J, Levi E, Phillip PA, Majumdar AP. Age-related increase in colorectal cancer stem cells in macroscopically normal mucosa of patients with adenomas: A risk factor for colon cancer. *Biochem Biophys Res Commun* 378 (3), 344-347 (2009).

[370] Pollard, T.D., Earnshaw, W.C., Cell Biology. *Elseiver, USA* (2004).

[371] Schubert, W., On the origin of cell functions encoded in the toponome. *J Biotechnol*.

[372] Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Böckelmann R, Malykh Y, Gollnick H, Friedenberger M, Bode M, Dress AW. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat Biotechnol* 24 (10), 1270-1278 (2006).

[373] Friedenberger, M., Bode, M., Krusche, A., & Schubert, W., Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures. *Nat Protoc* 2 (9), 2285-2294 (2007).

[374] Schubert, W., Multiple antigen—mapping microscopy of human tissue. *Excerpta Medica. Elsevier, Amsterdam. Adv. Anal. Cell. Pathol.*, 97–98 (1990).

[375] Schubert, W., Antigenic determinants of T lymphocyte alpha beta receptor and other leukocyte surface proteins as differential markers of skeletal muscle regeneration: detection of spatially and timely restricted patterns by MAM microscopy. *Eur J Cell Biol* 58 (2), 395-410 (1992).

[376] De Angelis L, Berghella L, Coletta M, Lattanzi L, Zanchi M, Cusella-De Angelis MG, Ponzetto C, Cossu G. Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. *J Cell Biol* 147 (4), 869-878 (1999).

[377] Schubert, W., Polymositis, topological proteomics technology and paradigm for cell invasion dynamics. *J. Theor. Med.* 4, 67–74. (2002).

[378] Abbott, A., Mapping togetherness (Research Highlight) (referring to Schubert et al. Nature 2006.) *Nature* 443, 609 (2006).

[379] Schubert, W., Gieseler, A., Krusche, A., & Hillert, R., Toponome mapping in prostate cancer: detection of 2000 cell surface protein clusters in a single tissue section and cell type specific annotation by using a three symbol code. *J Proteome Res* 8 (6), 2696-2707 (2009).

[380] Ecker, R.C. & Tarnok, A., Cytomics goes 3D: toward tissomics. *Cytometry A* 65 (1), 1-3 (2005).

[381] Ecker, R.C., Rogojanu, R., Streit, M., Oesterreicher, K., & Steiner, G.E., An improved method for discrimination of cell populations in tissue sections using microscopy-based multicolor tissue cytometry. *Cytometry A* 69 (3), 119-123 (2006).

[382] Mittag, A., Lenz, D., Gerstner, A.O., & Tarnok, A., Hyperchromatic cytometry principles for cytomics using slide based cytometry. *Cytometry A* 69 (7), 691-703 (2006).

[383] Laffers, W., Mittag, A., Lenz, D., Tarnok, A., & Gerstner, A.O., Iterative restaining as a pivotal tool for n-color immunophenotyping by slide-based cytometry. *Cytometry A* 69 (3), 127-130 (2006).

[384] Bode M, Irmler M, Friedenberger M, May C, Jung K, Stephan C, Meyer HE, Lach C, Hillert R, Krusche A, Beckers J, Marcus K, Schubert W. Interlocking transcriptomics, proteomics and toponomics technologies for brain tissue analysis in murine hippocampus. *Proteomics* 8 (6), 1170-1178 (2008).

[385] Bedner, E., Du, L., Traganos, F., & Darzynkiewicz, Z., Caffeine dissociates complexes between DNA and intercalating dyes: application for bleaching fluorochrome-stained cells for their subsequent restaining and analysis by laser scanning cytometry. *Cytometry* 43 (1), 38-45 (2001).

[386] Haars, R., Schneider, A., Bode, M., & Schubert, W., Secretion and differential localization of the proteolytic cleavage products Abeta40 and Abeta42 of the Alzheimer amyloid precursor protein in human fetal myogenic cells. *Eur J Cell Biol* 79 (6), 400-406 (2000).

[387] Ebert MP, Ademmer K, Müller-Ostermeyer F, Friess H, Büchler MW, Schubert W, Malfertheiner P. CD8+CD103+ T cells analogous to intestinal intraepithelial

lymphocytes infiltrate the pancreas in chronic pancreatitis. *Am J Gastroenterol* 93 (11), 2141-2147 (1998).

388 Mohamed HA, Mosier DR, Zou LL, Siklós L, Alexianu ME, Engelhardt JI, Beers DR, Le WD, Appel SH. Immunoglobulin Fc gamma receptor promotes immunoglobulin uptake, immunoglobulin-mediated calcium increase, and neurotransmitter release in motor neurons. *J Neurosci Res* 69 (1), 110-116 (2002).

389 Berndt U, Philipsen L, Bartsch S, Wiedenmann B, Baumgart DC, Hämmerle M, Sturm A. Systematic high-content proteomic analysis reveals substantial immunologic changes in colorectal cancer. *Cancer Res* 68 (3), 880-888 (2008).

390 Yue, S.Q., Yang, Y.L., Dou, K.F., & Li, K.Z., Expression of PCNA and CD44mRNA in colorectal cancer with venous invasion and its relationship to liver metastasis. *World J Gastroenterol* 9 (12), 2863-2865 (2003).

391 Bahnassy AA, Zekri AR, El-Houssini S, El-Shehaby AM, Mahmoud MR, Abdallah S, El-Serafi M. Cyclin A and cyclin D1 as significant prognostic markers in colorectal cancer patients. *BMC Gastroenterol* 4, 22 (2004).

392 Guo B, Zhai D, Cabezas E, Welsh K, Nouraini S, Satterthwait AC, Reed JC. Humanin peptide suppresses apoptosis by interfering with Bax activation. *Nature* 423 (6938), 456-461 (2003).

393 Pignatelli M, Liu D, Nasim MM, Stamp GW, Hirano S, Takeichi M. Morphoregulatory activities of E-cadherin and beta-1 integrins in colorectal tumour cells. *Br J Cancer* 66 (4), 629-634 (1992).

394 Niewiarowska J, Sacewicz I, Wiktorska M, Wysocki T, Stasikowska O, Wagrowska-Danilewicz M, Cierniewski CS. DNAzymes to mouse beta1 integrin mRNA in vivo: targeting the tumor vasculature and retarding cancer growth. *Cancer Gene Ther* 16 (9), 713-722 (2009).

[395] Hammarstrom, S., The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* 9 (2), 67-81 (1999).

[396] Blumenthal, R.D., Hansen, H.J., & Goldenberg, D.M., Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). *Cancer Res* 65 (19), 8809-8817 (2005).

[397] Tan E, Gouvas N, Nicholls RJ, Ziprin P, Xynos E, Tekkis PP. Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surg Oncol* 18 (1), 15-24 (2009).

[398] Li, Y.Y., Hsieh, L.L., Tang, R.P., Liao, S.K., & Yeh, K.Y., Macrophage-derived interleukin-6 up-regulates MUC1, but down-regulates MUC2 expression in the human colon cancer HT-29 cell line. *Cell Immunol* 256 (1-2), 19-26 (2009).

[399] Matsuda K, Masaki T, Watanabe T, Kitayama J, Nagawa H, Muto T, Ajioka Y. Clinical significance of MUC1 and MUC2 mucin and p53 protein expression in colorectal carcinoma. *Jpn J Clin Oncol* 30 (2), 89-94 (2000).

[400] Tanious, F.A., Veal, J.M., Buczak, H., Ratmeyer, L.S., & Wilson, W.D., DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* 31 (12), 3103-3112 (1992).

[401] Durbeej, M., Laminins. *Cell Tissue Res* (2009).

[402] Cho, S. & Kim, E., CD36: a multi-modal target for acute stroke therapy. *J Neurochem* 109 Suppl 1, 126-132 (2009).

[403] Togashi, H., Sakisaka, T., & Takai, Y., Cell adhesion molecules in the central nervous system. *Cell Adh Migr* 3 (1) (2009).

[404] Golby, S.J., Chinyama, C., & Spencer, J., Proliferation of T-cell subsets that contact tumour cells in colorectal cancer. *Clin Exp Immunol* 127 (1), 85-91 (2002).

[405] Needham, L.K. & Schnaar, R.L., The HNK-1 reactive sulfoglucuronyl glycolipids are ligands for L-selectin and P-selectin but not E-selectin. *Proc Natl Acad Sci U S A* 90 (4), 1359-1363 (1993).

[406] Uusitalo, M. & Kivela, T., The HNK-1 carbohydrate epitope in the eye: basic science and functional implications. *Prog Retin Eye Res* 20 (1), 1-28 (2001).

[407] Moll, R., Lowe, A., Laufer, J., & Franke, W.W., Cytokeratin 20 in human carcinomas. A new histodiagnostic marker detected by monoclonal antibodies. *Am J Pathol* 140 (2), 427-447 (1992).

[408] Hernandez BY, Frierson HF, Moskaluk CA, Li YJ, Clegg L, Cote TR, McCusker ME, Hankey BF, Edwards BK, Goodman MT. CK20 and CK7 protein expression in colorectal cancer: demonstration of the utility of a population-based tissue microarray. *Hum Pathol* 36 (3), 275-281 (2005).

[409] Lagoudianakis EE, Kataki A, Manouras A, Memos N, Papadima A, Derventzi A, Zografos G, Papadopoulos S, Katergiannakis V, Konstadoulakis MM. Detection of epithelial cells by RT-PCR targeting CEA, CK20, and TEM-8 in colorectal carcinoma patients using OncoQuick density gradient centrifugation system. *J Surg Res* 155 (2), 183-190 (2009).

[410] Dumay A, Laulier C, Bertrand P, Saintigny Y, Lebrun F, Vayssière JL, Lopez BS. Bax and Bid, two proapoptotic Bcl-2 family members, inhibit homologous recombination, independently of apoptosis regulation. *Oncogene* 25 (22), 3196-3205 (2006).

411 Andreadis D, Epivatianos A, Poulopoulos A, Nomikos A, Christidis K, Papazoglou G, Antoniades D, Barbatis C. Immunohistochemical detection of the expression of the cell adhesion molecules E-cadherin, desmoglein-2, beta4-integrin, ICAM-1 and HCAM (CD44s) in Warthin's tumour of the parotid gland. *Oral Oncol* 41 (8), 799-805 (2005).

412 Zimmer, R. & Thomas, P., Mutations in the carcinoembryonic antigen gene in colorectal cancer patients: implications on liver metastasis. *Cancer Res* 61 (7), 2822-2826 (2001).

413 Bhattacharya S, Mathew G, Ruban E, Epstein DB, Krusche A, Hillert R, Schubert W, Khan M. Toponome Imaging System: in situ protein network mapping in normal and cancerous colon from the same patient reveals more than five-thousand cancer specific protein clusters and their sub-cellular annotation by using a three symbol code. *J Proteome Res.* 2010 Dec 3;9(12):6112-25. Epub 2010 Oct 28

414 Yu, Y., Flint, A., Dvorin, E.L., & Bischoff, J., AC133-2, a novel isoform of human AC133 stem cell antigen. *J Biol Chem* 277 (23), 20711-20716 (2002).

415 Corbeil, D., Roper, K., Fargeas, C.A., Joester, A., & Huttner, W.B., Prominin: a story of cholesterol, plasma membrane protrusions and human pathology. *Traffic* 2 (2), 82-91 (2001).

416 Fargeas CA, Joester A, Missol-Kolka E, Hellwig A, Huttner WB, Corbeil D. Identification of novel Prominin-1/CD133 splice variants with alternative C-termini and their expression in epididymis and testis. *J Cell Sci* 117 (Pt 18), 4301-4311 (2004).

417 Marzesco AM, Janich P, Wilsch-Bräuninger M, Dubreuil V, Langenfeld K, Corbeil D, Huttner WB. Release of extracellular membrane particles carrying the

stem cell marker prominin-1 (CD133) from neural progenitors and other epithelial cells. *J Cell Sci* 118 (Pt 13), 2849-2858 (2005).

[418] Yin AH, Miraglia S, Zanjani ED, Almeida-Porada G, Ogawa M, Leary AG, Olweus J, Kearney J, Buck DW. AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood* 90 (12), 5002-5012 (1997).

[419] Mizrak, D., Brittan, M., & Alison, M.R., CD133: molecule of the moment. *J Pathol* 214 (1), 3-9 (2008).

[420] Burkert, J., Wright, N.A., & Alison, M.R., Stem cells and cancer: an intimate relationship. *J Pathol* 209 (3), 287-297 (2006).

[421] Todaro M, Alea MP, Di Stefano AB, Cammareri P, Vermeulen L, Iovino F, Tripodo C, Russo A, Gulotta G, Medema JP, Stassi G. Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. *Cell Stem Cell* 1 (4), 389-402 (2007).

[422] LaBarge, M.A. & Bissell, M.J., Is CD133 a marker of metastatic colon cancer stem cells? *J Clin Invest* 118 (6), 2021-2024 (2008).

[423] Nikolova T, Wu M, Brumbarov K, Alt R, Opitz H, Boheler KR, Cross M, Wobus AM. WNT-conditioned media differentially affect the proliferation and differentiation of cord blood-derived CD133+ cells in vitro. *Differentiation* 75 (2), 100-111 (2007).

[424] Clement, V., Sanchez, P., de Tribolet, N., Radovanovic, I., & Ruiz i Altaba, A., HEDGEHOG-GLI1 signaling regulates human glioma growth, cancer stem cell self-renewal, and tumorigenicity. *Curr Biol* 17 (2), 165-172 (2007).

[425] Fan X, Matsui W, Khaki L, Stearns D, Chun J, Li YM, Eberhart CG. Notch pathway inhibition depletes stem-like cells and blocks engraftment in embryonal brain tumors. *Cancer Res* 66 (15), 7445-7452 (2006).

[426] Piccirillo SG, Reynolds BA, Zanetti N, Lamorte G, Binda E, Broggi G, Brem H, Olivi A, Dimeco F, Vescovi AL. Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature* 444 (7120), 761-765 (2006).

[427] Kojima M, Ishii G, Atsumi N, Fujii S, Saito N, Ochiai A. Immunohistochemical detection of CD133 expression in colorectal cancer: a clinicopathological study. *Cancer Sci* 99 (8), 1578-1583 (2008).

[428] Horst, D., Kriegl, L., Engel, J., Kirchner, T., & Jung, A., CD133 expression is an independent prognostic marker for low survival in colorectal cancer. *Br J Cancer* 99 (8), 1285-1289 (2008).

[429] Horst D, Scheel SK, Liebmann S, Neumann J, Maatz S, Kirchner T, Jung A. The cancer stem cell marker CD133 has high prognostic impact but unknown functional relevance for the metastasis of human colon cancer. *J Pathol* (2009).

[430] Okegawa, T., Pong, R.C., Li, Y., & Hsieh, J.T., The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochim Pol* 51 (2), 445-457 (2004).

[431] Orian-Rousseau, V., CD44, a therapeutic target for metastasising tumours. *Eur J Cancer* 46 (7), 1271-1277.

[432] Bowen MA, Patel DD, Li X, Modrell B, Malacko AR, Wang WC, Marquardt H, Neubauer M, Pesando JM, Francke U, et al. Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. *J Exp Med* 181 (6), 2213-2220 (1995).

[433] Swart, G.W., Activated leukocyte cell adhesion molecule (CD166/ALCAM): developmental and mechanistic aspects of cell clustering and cell migration. *Eur J Cell Biol* 81 (6), 313-321 (2002).

[434] Weichert, W., Knosel, T., Bellach, J., Dietel, M., & Kristiansen, G., ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. *J Clin Pathol* 57 (11), 1160-1164 (2004).

[435] Trzpis, M., McLaughlin, P.M., de Leij, L.M., & Harmsen, M.C., Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *Am J Pathol* 171 (2), 386-395 (2007).

[436] Litvinov SV, Balzar M, Winter MJ, Bakker HA, Briaire-de Bruijn IH, Prins F, Fleuren GJ, Warnaar SO. Epithelial cell adhesion molecule (Ep-CAM) modulates cell-cell interactions mediated by classic cadherins. *J Cell Biol* 139 (5), 1337-1348 (1997).

[437] Münz M, Kieu C, Mack B, Schmitt B, Zeidler R, Gires O. The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. *Oncogene* 23 (34), 5748-5758 (2004).

[438] Osta WA, Chen Y, Mikhitarian K, Mitas M, Salem M, Hannun YA, Cole DJ, Gillanders WE. EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. *Cancer Res* 64 (16), 5818-5824 (2004).

[439] Maetzel D, Denzel S, Mack B, Canis M, Went P, Benk M, Kieu C, Papior P, Baeuerle PA, Munz M, Gires O. Nuclear signalling by tumour-associated antigen EpCAM. *Nat Cell Biol* 11 (2), 162-171 (2009).

[440] Carpenter, G. & Red Brewer, M., EpCAM: another surface-to-nucleus missile. *Cancer Cell* 15 (3), 165-166 (2009).

[441] Johannessen, M., Moller, S., Hansen, T., Moens, U., & Van Ghelue, M., The multifunctional roles of the four-and-a-half-LIM only protein FHL2. *Cell Mol Life Sci* 63 (3), 268-284 (2006).

[442] Baeuerle, P.A. & Gires, O., EpCAM (CD326) finding its role in cancer. *Br J Cancer* 96 (3), 417-423 (2007).

[443] Ponta, H., Sherman, L., & Herrlich, P.A., CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol* 4 (1), 33-45 (2003).

[444] Pure, E. & Assoian, R.K., Rheostatic signaling by CD44 and hyaluronan. *Cell Signal* 21 (5), 651-655 (2009).

[445] Jones, M., Tussey, L., Athanasou, N., & Jackson, D.G., Heparan sulfate proteoglycan isoforms of the CD44 hyaluronan receptor induced in human inflammatory macrophages can function as paracrine regulators of fibroblast growth factor action. *J Biol Chem* 275 (11), 7964-7974 (2000).

[446] Tsukita, S., Oishi, K., Sato, N., Sagara, J., & Kawai, A., ERM family members as molecular linkers between the cell surface glycoprotein CD44 and actin-based cytoskeletons. *J Cell Biol* 126 (2), 391-401 (1994).

[447] Legg, J.W., Lewis, C.A., Parsons, M., Ng, T., & Isacke, C.M., A novel PKC-regulated mechanism controls CD44 ezrin association and directional cell motility. *Nat Cell Biol* 4 (6), 399-407 (2002).

[448] Günthert U, Hofmann M, Rudy W, Reber S, Zöller M, Haussmann I, Matzku S, Wenzel A, Ponta H, Herrlich P. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell* 65 (1), 13-24 (1991).

[449] Seiter S, Arch R, Reber S, Komitowski D, Hofmann M, Ponta H, Herrlich P, Matzku S, Zöller M. Prevention of tumor metastasis formation by anti-variant CD44. *J Exp Med* 177 (2), 443-455 (1993).

[450] Wielenga, V.J., van der Neut, R., Offerhaus, G.J., & Pals, S.T., CD44 glycoproteins in colorectal cancer: expression, function, and prognostic value. *Adv Cancer Res* 77, 169-187 (2000).

[451] Yu, Q., Toole, B.P., & Stamenkovic, I., Induction of apoptosis of metastatic mammary carcinoma cells in vivo by disruption of tumor cell surface CD44 function. *J Exp Med* 186 (12), 1985-1996 (1997).

[452] Yu, Q. & Stamenkovic, I., Localization of matrix metalloproteinase 9 to the cell surface provides a mechanism for CD44-mediated tumor invasion. *Genes Dev* 13 (1), 35-48 (1999).

[453] Kim HR, Wheeler MA, Wilson CM, Iida J, Eng D, Simpson MA, McCarthy JB, Bullard KM. Hyaluronan facilitates invasion of colon carcinoma cells in vitro via interaction with CD44. *Cancer Res* 64 (13), 4569-4576 (2004).

[454] Hsieh, H.F., Yu, J.C., Ho, L.I., Chiu, S.C., & Harn, H.J., Molecular studies into the role of CD44 variants in metastasis in gastric cancer. *Mol Pathol* 52 (1), 25-28 (1999).

[455] Gotte, M. & Yip, G.W., Heparanase, hyaluronan, and CD44 in cancers: a breast carcinoma perspective. *Cancer Res* 66 (21), 10233-10237 (2006).

[456] Naor, D., Sionov, R.V., & Ish-Shalom, D., CD44: structure, function, and association with the malignant process. *Adv Cancer Res* 71, 241-319 (1997).

[457] Weber, G.F., The metastasis gene osteopontin: a candidate target for cancer therapy. *Biochim Biophys Acta* 1552 (2), 61-85 (2001).

[458] Takafuji, V., Forgues, M., Unsworth, E., Goldsmith, P., & Wang, X.W., An osteopontin fragment is essential for tumor cell invasion in hepatocellular carcinoma. *Oncogene* 26 (44), 6361-6371 (2007).

[459] Tijink BM, Buter J, de Bree R, Giaccone G, Lang MS, Staab A, Leemans CR, van Dongen GA. A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients with incurable squamous cell carcinoma of the head and neck or esophagus. *Clin Cancer Res* 12 (20 Pt 1), 6064-6072 (2006).

[460] Riechelmann H, Sauter A, Golze W, Hanft G, Schroen C, Hoermann K, Erhardt T, Gronau S. Phase I trial with the CD44v6-targeting immunoconjugate bivatuzumab mertansine in head and neck squamous cell carcinoma. *Oral Oncol* 44 (9), 823-829 (2008).

[461] Fuchs, E. & Cleveland, D.W., A structural scaffolding of intermediate filaments in health and disease. *Science* 279 (5350), 514-519 (1998).

[462] Schweizer J, Bowden PE, Coulombe PA, Langbein L, Lane EB, Magin TM, Maltais L, Omary MB, Parry DA, Rogers MA, Wright MW. New consensus nomenclature for mammalian keratins. *J Cell Biol* 174 (2), 169-174 (2006).

[463] Ku, N.O., Zhou, X., Toivola, D.M., & Omary, M.B., The cytoskeleton of digestive epithelia in health and disease. *Am J Physiol* 277 (6 Pt 1), G1108-1137 (1999).

[464] Herrmann, H., Strelkov, S.V., Burkhard, P., & Aebi, U., Intermediate filaments: primary determinants of cell architecture and plasticity. *J Clin Invest* 119 (7), 1772-1783 (2009).

[465] Moll, R., Molecular diversity of cytokeratins: significance for cell and tumor differentiation. *Acta Histochem Suppl* 41, 117-127 (1991).

[466] Toivola, D.M., Tao, G.Z., Habtezion, A., Liao, J., & Omary, M.B., Cellular integrity plus: organelle-related and protein-targeting functions of intermediate filaments. *Trends Cell Biol* 15 (11), 608-617 (2005).

[467] Oriolo, A.S., Wald, F.A., Ramsauer, V.P., & Salas, P.J., Intermediate filaments: a role in epithelial polarity. *Exp Cell Res* 313 (10), 2255-2264 (2007).

[468] Ku NO, Toivola DM, Zhou Q, Tao GZ, Zhong B, Omary MB. Studying simple epithelial keratins in cells and tissues. *Methods Cell Biol* 78, 489-517 (2004).

[469] Moll, R., Divo, M., & Langbein, L., The human keratins: biology and pathology. *Histochem Cell Biol* 129 (6), 705-733 (2008).

[470] Ku, N.O., Strnad, P., Zhong, B.H., Tao, G.Z., & Omary, M.B., Keratins let liver live: Mutations predispose to liver disease and crosslinking generates Mallory-Denk bodies. *Hepatology* 46 (5), 1639-1649 (2007).

[471] Birkenkamp-Demtroder K, Mansilla F, Sørensen FB, Kruhøffer M, Cabezón T, Christensen LL, Aaltonen LA, Verspaget HW, Ørntoft TF. Phosphoprotein Keratin 23 accumulates in MSS but not MSI colon cancers in vivo and impacts viability and proliferation in vitro. *Mol Oncol* 1 (2), 181-195 (2007).

[472] Rogers, M.A., Winter, H., Langbein, L., Bleiler, R., & Schweizer, J., The human type I keratin gene family: characterization of new hair follicle specific members and evaluation of the chromosome 17q21.2 gene domain. *Differentiation* 72 (9-10), 527-540 (2004).

[473] Zhang, J.S., Wang, L., Huang, H., Nelson, M., & Smith, D.I., Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. *Genes Chromosomes Cancer* 30 (2), 123-135 (2001).

[474] Polley AC, Mulholland F, Pin C, Williams EA, Bradburn DM, Mills SJ, Mathers JC, Johnson IT. Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. *Cancer Res* 66 (13), 6553-6562 (2006).

[475] Toivola, D.M., Baribault, H., Magin, T., Michie, S.A., & Omary, M.B., Simple epithelial keratins are dispensable for cytoprotection in two pancreatitis models. *Am J Physiol Gastrointest Liver Physiol* 279 (6), G1343-1354 (2000).

[476] Fradette, J., Germain, L., Seshaiah, P., & Coulombe, P.A., The type I keratin 19 possesses distinct and context-dependent assembly properties. *J Biol Chem* 273 (52), 35176-35184 (1998).

[477] Zhou Q, Cadrin M, Herrmann H, Chen CH, Chalkley RJ, Burlingame AL, Omary MB. Keratin 20 serine 13 phosphorylation is a stress and intestinal goblet cell marker. *J Biol Chem* 281 (24), 16453-16461 (2006).

[478] Chu, P.G. & Weiss, L.M., Keratin expression in human tissues and neoplasms. *Histopathology* 40 (5), 403-439 (2002).

[479] Tot, T., Cytokeratins 20 and 7 as biomarkers: usefulness in discriminating primary from metastatic adenocarcinoma. *Eur J Cancer* 38 (6), 758-763 (2002).

[480] McGregor, D.K., Wu, T.T., Rashid, A., Luthra, R., & Hamilton, S.R., Reduced expression of cytokeratin 20 in colorectal carcinomas with high levels of microsatellite instability. *Am J Surg Pathol* 28 (6), 712-718 (2004).

[481] Woelfle, U., Sauter, G., Santjer, S., Brakenhoff, R., & Pantel, K., Down-regulated expression of cytokeratin 18 promotes progression of human breast cancer. *Clin Cancer Res* 10 (8), 2670-2674 (2004).

[482] Buhler, H. & Schaller, G., Transfection of keratin 18 gene in human breast cancer cells causes induction of adhesion proteins and dramatic regression of malignancy in vitro and in vivo. *Mol Cancer Res* 3 (7), 365-371 (2005).

[483] Omary, M.B., Ku, N.O., Strnad, P., & Hanada, S., Toward unraveling the complexity of simple epithelial keratins in human disease. *J Clin Invest* 119 (7), 1794-1805 (2009).

[484] Hollingsworth, M.A. & Swanson, B.J., Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer* 4 (1), 45-60 (2004).

[485] Moniaux, N., Escande, F., Porchet, N., Aubert, J.P., & Batra, S.K., Structural organization and classification of the human mucin genes. *Front Biosci* 6, D1192-1206 (2001).

[486] Taylor-Papadimitriou, J., Burchell, J., Miles, D.W., & Dalziel, M., MUC1 and cancer. *Biochim Biophys Acta* 1455 (2-3), 301-313 (1999).

[487] Moniaux, N., Junker, W.M., Singh, A.P., Jones, A.M., & Batra, S.K., Characterization of human mucin MUC17. Complete coding sequence and organization. *J Biol Chem* 281 (33), 23676-23685 (2006).

[488] Croce, C.M., Oncogenes and cancer. *N Engl J Med* 358 (5), 502-511 (2008).

[489] Li Y, Ren J, Yu W, Li Q, Kuwahara H, Yin L, Carraway KL 3rd, Kufe D. The epidermal growth factor receptor regulates interaction of the human DF3/MUC1 carcinoma antigen with c-Src and beta-catenin. *J Biol Chem* 276 (38), 35239-35242 (2001).

[490] Singh, A.P., Moniaux, N., Chauhan, S.C., Meza, J.L., & Batra, S.K., Inhibition of MUC4 expression suppresses pancreatic tumor cell growth and metastasis. *Cancer Res* 64 (2), 622-630 (2004).

[491] Bafna S, Singh AP, Moniaux N, Eudy JD, Meza JL, Batra SK. MUC4, a multifunctional transmembrane glycoprotein, induces oncogenic transformation of NIH3T3 mouse fibroblast cells. *Cancer Res* 68 (22), 9231-9238 (2008).

[492] Chaturvedi P, Singh AP, Moniaux N, Senapati S, Chakraborty S, Meza JL, Batra SK. MUC4 mucin potentiates pancreatic tumor cell proliferation, survival, and invasive properties and interferes with its interaction to extracellular matrix proteins. *Mol Cancer Res* 5 (4), 309-320 (2007).

[493] Moniaux N, Chaturvedi P, Varshney GC, Meza JL, Rodriguez-Sierra JF, Aubert JP, Batra SK. Human MUC4 mucin induces ultra-structural changes and tumorigenicity in pancreatic cancer cells. *Br J Cancer* 97 (3), 345-357 (2007).

[494] Gendler, S.J., MUC1, the renaissance molecule. *J Mammary Gland Biol Neoplasia* 6 (3), 339-353 (2001).

[495] Osako M, Yonezawa S, Siddiki B, Huang J, Ho JJ, Kim YS, Sato E. Immunohistochemical study of mucin carbohydrates and core proteins in human pancreatic tumors. *Cancer* 71 (7), 2191-2199 (1993).

[496] Yonezawa, S. & Sato, E., Expression of mucin antigens in human cancers and its relationship with malignancy potential. *Pathol Int* 47 (12), 813-830 (1997).

[497] Yamashita K, Yonezawa S, Tanaka S, Shirahama H, Sakoda K, Imai K, Xing PX, McKenzie IF, Hilkens J, Kim YS, *et al.* Immunohistochemical study of mucin carbohydrates and core proteins in hepatolithiasis and cholangiocarcinoma. *Int J Cancer* 55 (1), 82-91 (1993).

[498] Kitamura, H., Yonezawa, S., Tanaka, S., Kim, Y.S., & Sato, E., Expression of mucin carbohydrates and core proteins in carcinomas of the ampulla of Vater: their relationship to prognosis. *Jpn J Cancer Res* 87 (6), 631-640 (1996).

[499] Yonezawa S, Sueyoshi K, Nomoto M, Kitamura H, Nagata K, Arimura Y, Tanaka S, Hollingsworth MA, Siddiki B, Kim YS, Sato E. MUC2 gene expression is found in noninvasive tumors but not in invasive tumors of the pancreas and liver: its close relationship with prognosis of the patients. *Hum Pathol* 28 (3), 344-352 (1997).

[500] Utsunomiya T, Yonezawa S, Sakamoto H, Kitamura H, Hokita S, Aiko T, Tanaka S, Irimura T, Kim YS, Sato E. Expression of MUC1 and MUC2 mucins in gastric

carcinomas: its relationship with the prognosis of the patients. *Clin Cancer Res* 4 (11), 2605-2614 (1998).

501 Higashi M, Yonezawa S, Ho JJ, Tanaka S, Irimura T, Kim YS, Sato E. Expression of MUC1 and MUC2 mucin antigens in intrahepatic bile duct tumors: its relationship with a new morphological classification of cholangiocarcinoma. *Hepatology* 30 (6), 1347-1355 (1999).

502 Yonezawa, S., Nakamura, A., Horinouchi, M., & Sato, E., The expression of several types of mucin is related to the biological behavior of pancreatic neoplasms. *J Hepatobiliary Pancreat Surg* 9 (3), 328-341 (2002).

503 Nakamura A, Horinouchi M, Goto M, Nagata K, Sakoda K, Takao S, Imai K, Kim YS, Sato E, Yonezawa S. New classification of pancreatic intraductal papillary-mucinous tumour by mucin expression: its relationship with potential for malignancy. *J Pathol* 197 (2), 201-210 (2002).

504 Tamada S, Goto M, Nomoto M, Nagata K, Shimizu T, Tanaka S, Sakoda K, Imai K, Yonezawa S. Expression of MUC1 and MUC2 mucins in extrahepatic bile duct carcinomas: its relationship with tumor progression and prognosis. *Pathol Int* 52 (11), 713-723 (2002).

505 Shibahara H, Tamada S, Higashi M, Goto M, Batra SK, Hollingsworth MA, Imai K, Yonezawa S. MUC4 is a novel prognostic factor of intrahepatic cholangiocarcinoma-mass forming type. *Hepatology* 39 (1), 220-229 (2004).

506 Saitou M, Goto M, Horinouchi M, Tamada S, Nagata K, Hamada T, Osako M, Takao S, Batra SK, Aikou T, Imai K, Yonezawa S. MUC4 expression is a novel prognostic factor in patients with invasive ductal carcinoma of the pancreas. *J Clin Pathol* 58 (8), 845-852 (2005).

[507] Tsutsumida H, Goto M, Kitajima S, Kubota I, Hirotsu Y, Wakimoto J, Batra SK, Imai K, Yonezawa S. MUC4 expression correlates with poor prognosis in small-sized lung adenocarcinoma. *Lung Cancer* 55 (2), 195-203 (2007).

[508] Hattrup, C.L. & Gendler, S.J., Structure and function of the cell surface (tethered) mucins. *Annu Rev Physiol* 70, 431-457 (2008).

[509] Yonezawa, S., Application of immunohistochemistry for diagnosis of neoplasms: Mucin antigens expression and biological behavior. *Acta Histochem. Cytochem.* 1994, (*27*,), 561-566. (1994).

[510] Zhao Q, Guo X, Nash GB, Stone PC, Hilkens J, Rhodes JM, Yu LG. Circulating galectin-3 promotes metastasis by modifying MUC1 localization on cancer cell surface. *Cancer Res* 69 (17), 6799-6806 (2009).

[511] Singh, P.K. & Hollingsworth, M.A., Cell surface-associated mucins in signal transduction. *Trends Cell Biol* 16 (9), 467-476 (2006).

[512] Lan, M.S., Batra, S.K., Qi, W.N., Metzgar, R.S., & Hollingsworth, M.A., Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J Biol Chem* 265 (25), 15294-15299 (1990).

[513] Pochampalli, M.R., el Bejjani, R.M., & Schroeder, J.A., MUC1 is a novel regulator of ErbB1 receptor trafficking. *Oncogene* 26 (12), 1693-1701 (2007).

[514] Schroeder, J.A., Thompson, M.C., Gardner, M.M., & Gendler, S.J., Transgenic MUC1 interacts with epidermal growth factor receptor and correlates with mitogen-activated protein kinase activation in the mouse mammary gland. *J Biol Chem* 276 (16), 13057-13064 (2001).

[515] Hattrup, C.L. & Gendler, S.J., MUC1 alters oncogenic events and transcription in human breast cancer cells. *Breast Cancer Res* 8 (4), R37 (2006).

[516] Ren J, Raina D, Chen W, Li G, Huang L, Kufe D. MUC1 oncoprotein functions in activation of fibroblast growth factor receptor signaling. *Mol Cancer Res* 4 (11), 873-883 (2006).

[517] Huang L, Chen D, Liu D, Yin L, Kharbanda S, Kufe D. MUC1 oncoprotein blocks glycogen synthase kinase 3beta-mediated phosphorylation and degradation of beta-catenin. *Cancer Res* 65 (22), 10413-10422 (2005).

[518] Wei, X., Xu, H., & Kufe, D., MUC1 oncoprotein stabilizes and activates estrogen receptor alpha. *Mol Cell* 21 (2), 295-305 (2006).

[519] Wei, X., Xu, H., & Kufe, D., Human MUC1 oncoprotein regulates p53-responsive gene transcription in the genotoxic stress response. *Cancer Cell* 7 (2), 167-178 (2005).

[520] Raina, D., Kharbanda, S., & Kufe, D., The MUC1 oncoprotein activates the anti-apoptotic phosphoinositide 3-kinase/Akt and Bcl-xL pathways in rat 3Y1 fibroblasts. *J Biol Chem* 279 (20), 20607-20612 (2004).

[521] Yin, L., Li, Y., Ren, J., Kuwahara, H., & Kufe, D., Human MUC1 carcinoma antigen regulates intracellular oxidant levels and the apoptotic response to oxidative stress. *J Biol Chem* 278 (37), 35458-35464 (2003).

[522] Kharbanda S, Pandey P, Ren R, Mayer B, Zon L, Kufe D. c-Abl activation regulates induction of the SEK1/stress-activated protein kinase pathway in the cellular response to 1-beta-D-arabinofuranosylcytosine. *J Biol Chem* 270 (51), 30278-30281 (1995).

[523] Kharbanda S, Ren R, Pandey P, Shafman TD, Feller SM, Weichselbaum RR, Kufe DW. Activation of the c-Abl tyrosine kinase in the stress response to DNA-damaging agents. *Nature* 376 (6543), 785-788 (1995).

[524] Raina D, Ahmad R, Kumar S, Ren J, Yoshida K, Kharbanda S, Kufe D. MUC1 oncoprotein blocks nuclear targeting of c-Abl in the apoptotic response to DNA damage. *EMBO J* 25 (16), 3774-3783 (2006).

[525] Ahmad R, Raina D, Trivedi V, Ren J, Rajabi H, Kharbanda S, Kufe D. MUC1 oncoprotein activates the IkappaB kinase beta complex and constitutive NF-kappaB signalling. *Nat Cell Biol* 9 (12), 1419-1427 (2007).

[526] Agata N, Ahmad R, Kawano T, Raina D, Kharbanda S, Kufe D. MUC1 oncoprotein blocks death receptor-mediated apoptosis by inhibiting recruitment of caspase-8. *Cancer Res* 68 (15), 6136-6144 (2008).

[527] Bafna, S., Kaur, S., & Batra, S.K., Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene* 29 (20), 2893-2904.

[528] Blumenthal, R.D., Leon, E., Hansen, H.J., & Goldenberg, D.M., Expression patterns of CEACAM5 and CEACAM6 in primary and metastatic cancers. *BMC Cancer* 7, 2 (2007).

[529] Boman, B.M. & Huang, E., Human colon cancer stem cells: a new paradigm in gastrointestinal oncology. *J Clin Oncol* 26 (17), 2828-2838 (2008).

[530] Chu P, Clanton DJ, Snipas TS, Lee J, Mitchell E, Nguyen ML, Hare E, Peach RJ. Characterization of a subpopulation of colon cancer cells with stem cell-like properties. *Int J Cancer* (2008).

[531] Horst, D., Kriegl, L., Engel, J., Kirchner, T., & Jung, A., Prognostic significance of the cancer stem cell markers CD133, CD44, and CD166 in colorectal cancer. *Cancer Invest* 27 (8), 844-850 (2009).

[532] Neuzil J, Stantic M, Zobalova R, Chladova J, Wang X, Prochazka L, Dong L, Andera L, Ralph SJ. Tumour-initiating cells vs. cancer 'stem' cells and CD133: what's in the name? *Biochem Biophys Res Commun* 355 (4), 855-859 (2007).

[533] Siegel CA, Schwartz LM, Woloshin S, Cole EB, Rubin DT, Vay T, Baars J, Sands BE. When should ulcerative colitis patients undergo colectomy for dysplasia? Mismatch between patient preferences and physician recommendations. *Inflamm Bowel Dis*.

Table 1: MELK/TIS tag library.

Molecule/ Moiety recognised	Official symbol	Cellular expression / Known Function	Position in TIS
CD44	**CD44**	CD44 or HCAM (Homing cell adhesion molecule) is a cell-surface glycoprotein involved in cell-cell interactions, cell adhesion and migration. This protein participates in a wide variety of cellular functions including lymphocyte activation, recirculation and homing, hematopoiesis, and tumor metastasis. [*Homo sapiens*]	**4**
Cyclin D1	**CCND1**	This cyclin forms a complex with and functions as a regulatory subunit of CDK4 or CDK6 (CDK Kinase), whose activity is required for cell cycle G1/S transition. Mutations, amplification and overexpression of this gene, which alters cell cycle progression, are observed frequently in a variety of tumors and may contribute to tumorigenesis. [*Homo sapiens*]	**5**
Cytokeratin-20	**KRT20**	The keratins are intermediate filament proteins responsible for the structural integrity of epithelial cells. This cytokeratin is a major cellular protein of mature enterocytes and goblet cells and is specifically expressed in the gastric and intestinal mucosa. It is known to be over-expressed in colon adenocarcinoma as compared to normal colon. [*Homo sapiens*]	**6**
EpCAM	**EpCAM**	Epithelial cell adhesion molecule is expressed on most normal epithelial cells and gastrointestinal carcinomas and functions as a homotypic calcium-independent cell adhesion molecule. The antigen is being used as a target for immunotherapy treatment of human carcinomas.[*Homo sapiens*]	**7**
Cyclin A	**CCNA2**	This cyclin is expressed in all tissues tested. It binds and activates CDC2 or CDK2 kinases, and thus promotes both cell cycle G1/S and G2/M transitions. [*Homo sapiens*]	**8**

CD166	**ALCAM**	Activated leukocyte cell adhesion molecule is a trans-membrane glycoprotein. The main functions of this protein are ligand binding, cellular avidity and development. It is found in selected epithelial cells, lymphoid and myeloid cells, bone marrow mesenchymal cells, hepatocytes, pancreatic islets and acinar cells, developing embryo, several cancers like colorectal, breast, lung, prostrate and melanomas. Its degree of induction co-relates with the invasiveness of the respective cancers. [*Homo sapiens*]	**9**
Cytokeratin-19	**KRT19**	This smallest known acidic cytokeratin is not paired with a basic cytokeratin in epithelial cells. It is specifically expressed in the periderm, the transiently superficial layer that envelopes the developing epidermis. [*Homo sapiens*]	**10**
Ki-67	**MKI67**	It is a nuclear protein that is associated with and may be necessary for cellular proliferation. [*Homo sapiens*]	**11**
Mucin-1	**MUC3A**	It is an epithelial glycoprotein, which is secreted and some of it is membrane bound. Each of its genes contain at least one large domain of tandemly repeated sequence that encodes the peptide sequence rich in serine and/or threonine residues, which carries most of the O-linked glycosylation. [*Homo sapiens*]	**12**
Carcino Embryonic Antigen	**CEA**	This is a glycoprotein, with a series of Ig like domains. Its normal tissue distribution includes columnar epithelial cells and goblet cells in colon, mucous cells in stomach, squamous epithelium of tongue, esophagus and cervix and prostate. It is used clinically as a tumour marker for colorectal cancer. It is also elevated in cirrhosis, biliary obstruction and hepatitis. Highest levels are encountered in hepatic metastases from colon cancer. [*Homo sapiens*]	**13**

CD133	**PROM1**	Prominin-1 is a pentaspan transmembrane glycoprotein. The protein localizes to membrane protrusions and is often expressed on adult stem cells, where it is thought to function in maintaining stem cell properties by suppressing differentiation. [*Homo sapiens*]	**14**
Mucin-2	**MUC2**	It forms an insoluble mucous barrier that protects the gut lumen. The protein polymerizes into a gel of which 80% is composed of oligosaccharide side chains by weight. The protein features a central domain containing tandem repeats rich in threonine and proline that varies between 50 and 115 copies in different individuals. [*Homo sapiens*]	**15**
Laminin	**LAMB1**	It is an extracellular matrix glycoprotein and the major noncollagenous constituent of basement membranes. It is implicated in wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis. [*Homo sapiens*]	**17**
PCNA	**PCNA**	Proliferating cell nuclear antigen is found in the nucleus and is a cofactor of DNA polymerase delta. It acts as a homotrimer and promotes DNA replication. In response to DNA damage, this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway. [*Homo sapiens*]	**18**
Bcl-2	**BCL2**	It is an integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells such as lymphocytes. It is over expressed in various cancers. [*Homo sapiens*]	**19**
NCAM	**NCAM1**	Neural cell adhesion molecule is a major cell adhesion molecule found in several differentiated cells. It mediates adhesion between neurones and also between neurones and muscles. [*Homo sapiens*]	**20**

Bax	**BAX**	This protein functions as an apoptotic activator. It interacts with, and increases the opening of, the mitochondrial voltage-dependent anion channel (VDAC), which leads to the release of cytochrome c. The expression of this gene is regulated by the tumor suppressor P53 and has been shown to be involved in P53-mediated apoptosis. [*Homo sapiens*]	**21**
CD29	**ITGB1**	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) is expressed by all leukocytes and forms heterodimers with many integrin R subunits which mediate cell-cell and cell-matrix adhesion. [*Homo sapiens*]	**22**
CD36	**CD36**	It is a major glycoprotein on the platelet surface and serves as a receptor for thrombospondin in platelets and various cell lines. It binds to collagen, thrombospondin, anionic phospholipids and oxidized LDL. [*Homo sapiens*]	**23**
CD57	**CD57**	This glycoprotein is expressed normally in haematopoietic cells (Natural killer cells or NK cells and CD8 positive T lymphocytes), neuro-ectodermal cells, neuro endocrine cells, striated muscles and epithelium of prostate. It is possibly related to cell-cell interaction. Laminin, P-selectin and N-selectin are its natural ligands. Binding to the last two initiates rolling of leucocytes on endothelium and activated platelets. [*Homo sapiens*]	**24**
4,6’-diamidino-2-phenylindole	**DAPI**	4,6’-diamidino-2-phenylindole (DAPI) is an intercalating agent. It is a fluorescent molecule and has been used to stain DNA nucleic acids.	**25**

List of molecular cell components labelled in this study (column 1) together with their locus link annotations (column 2 and 3) and their respective positions in the cycle (column 4). Column 3 describes the known cellular expressions and functions of the corresponding protein markers, illustrating the rationale behind their selection in this study. Note that all the protein tags were conjugated to FITC in our experiments. DAPI was used on its own at the end of the experiment to tag the nuclei.

Table 2

PCNA	Bax	Bcl2	CD29	CD36	CD44	CD57	CD133	CD166	CEA	CK19	CK20	CyclinA	CyclinD1	DAPI	EpCAM	Ki67	Laminin	Muc2	Muc1	NCAM	Freq.	Colour codes
1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	469	
1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	382	
1	1	1	0	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	0	158	
1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	114	
1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	102	
1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	92	
1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	89	
1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	69	
1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	58	
1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	52	
1	1	1	0	0	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	0	47	
1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	28	
1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	27	
1	1	1	1	0	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	22	
1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	20	
1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	
1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	15	
1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	15	
1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	14	
1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	11	
1	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	10	
1	1	1	0	1	0	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	8	
1	1	1	1	0	0	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	6	
1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	5	
1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	5	
1	1	1	0	0	0	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	5	
1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	4	
1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	3	
1	1	1	0	0	0	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	3	
1	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	3	
1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	2	
1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	2	
1	1	1	0	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	
1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	

Table 3

PCNA	Bax	Bcl2	CD29	CD36	CD44	CD57	CD133	CD166	CEA	CK19	CK20	CyclinA	CyclinD1	DAPI	EpCAM	Ki67	Laminin	Muc2	Muc1	NCAM	Freq.	Colour codes
1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	5012	
1	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	2387	
1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	2187	
1	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	2057	
1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1787	
1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1738	
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1228	
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1189	
1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1075	
1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1049	
1	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	1049	
1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	988	
1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	918	
1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	762	
1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	728	
1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	705	
1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	1	0	0	0	658	
1	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	0	0	0	529	
1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	468	
1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	459	
1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	0	0	0	561	
1	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	417	
1	1	1	0	0	0	1	1	1	0	0	1	0	0	0	0	1	1	0	0	0	405	
1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	396	
1	0	1	0	0	0	1	1	0	0	0	1	0	0	1	0	1	1	0	0	0	392	

Table 4

PCNA	Bax	Bcl2	CD29	CD36	CD44	CD57	CD133	CD166	CEA	CK19	CK20	CyclinA	CyclinD1	DAPI	EpCAM	Ki67	Laminin	Muc2	Muc1	NCAM	Freq.	Colour Codes
1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1230	
1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1200	
0	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	492	
0	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	479	
1	1	1	0	0	0	1	0	0	1	1	1	0	0	0	1	1	1	1	1	0	469	
1	1	1	1	0	0	1	0	1	1	0	1	1	1	0	1	1	1	1	1	1	446	
1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	425	
1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	362	
1	1	1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	231	
0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	0	209	
1	1	1	0	0	0	1	0	1	1	1	1	0	0	0	1	1	1	0	1	0	192	
1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1	0	1	0	183	
1	1	1	0	0	0	1	0	0	1	1	1	0	0	0	1	1	1	0	1	0	182	
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	0	0	1	0	139	
1	1	1	0	0	0	1	0	0	1	0	1	0	0	0	1	1	1	0	1	0	130	
1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	117	
0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	1	0	0	1	0	114	
1	1	1	0	0	0	1	0	1	1	0	1	0	0	0	1	1	1	0	1	0	113	
0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	1	1	0	0	0	0	110	
1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0	104	
1	1	1	0	0	0	1	0	1	1	1	1	0	0	0	1	1	1	0	0	0	104	
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	0	0	0	0	102	
0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	0	0	0	0	101	
0	0	1	0	0	0	1	0	1	1	0	1	0	0	0	1	1	0	0	0	0	101	
0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0	92	
0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0	92	
0	0	1	0	0	0	1	0	1	1	1	1	0	0	0	1	1	1	0	0	0	92	
0	0	0	0	0	0	1	0	1	1	1	1	0	0	0	1	1	0	0	0	0	89	
0	1	0	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0	86	

A